

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:44:16 ; Search time 44 Seconds
(Without alignments) 345,210 Million cell updates/sec

Title: US-09-743-533-19
Perfect score: 906
Sequence: 1 MROLNFCSELOSPQOSYLO.....PPQOPQPVPOQASCIWSMV 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.73: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	830.5	91.7	260	2	seed storage prote
2	752	83.0	310	2	C-hordein precursor
3	580	64.0	347	2	probable hordein C
4	566	62.5	357	2	omega secalin prec
5	562.5	62.1	357	2	gamma secalin prec
6	412	45.5	327	2	gamma-gliadin prec
7	384.5	42.4	302	2	gamma-gliadin prec
8	347	38.3	251	2	gamma-gliadin prec
9	332	36.6	291	1	gamma-gliadin prec
10	310.5	34.3	105	2	gamma-gliadin B pr
11	302	33.3	305	2	gamma-gliadin (clone p
12	289.5	32.0	307	2	gamma-hordein 1 pr
13	279.5	30.8	509	2	alpha-beta-gliadin
14	272.5	30.1	72	2	egg envelope prote
15	271.5	30.0	293	2	C-hordein (pc horl
16	268.5	29.6	326	2	hordein B1 - barley
17	263.5	29.1	297	2	alpha-beta-gliadin
18	262.5	29.0	320	2	alpha-beta-gliadin
19	260.5	28.8	320	2	hordein B precursor
20	256.5	28.3	286	2	alpha-beta-gliadin
21	255.5	28.2	319	2	gamma-hordein 3 -
22	255	28.1	296	2	alpha-beta-gliadin
23	253.5	28.0	296	2	alpha-beta-gliadin
24	250	27.6	291	2	gliadin - wheat
25	244.5	27.0	1343	2	alpha/beta-gliadin
26	242.5	26.8	374	2	cell division prot
27	242	26.7	282	2	glutenin low molec
28	240	26.5	264	2	alpha/beta-gliadin
29	239	26.4	313	2	B3-hordein (clone
					alpha/beta-gliadin

30	238	26.3	356	2	S01992	glutenin low molec
31	238	26.3	359	2	T06982	glutenin low molec
32	237	26.2	288	2	T06282	alpha-gliadin prec
33	236.5	26.1	319	2	A22364	alpha/beta-gliadin
34	236	26.0	286	2	S07923	alpha/beta-gliadin
35	235.5	26.0	276	2	S57636	glutenin low molec
36	234	25.8	286	1	EWMTA	glutenin low molec
37	234	25.8	292	2	B22364	alpha/beta-gliadin
38	234	25.8	1342	2	G90750	alpha/beta-gliadin
39	234	25.8	1342	2	E85614	cell division prot
40	230.5	25.4	261	2	S57655	glutenin low molec
41	219	24.2	271	2	T04474	B1 hordein - barley
42	218.5	24.1	68	2	B25677	C-hordein (pc-919)
43	217	24.0	1329	2	A64828	cell division prot
44	211.5	23.3	307	2	S04325	glutenin low molec
45	208.5	23.0	462	1	LDD07	annexin VII - slim

ALIGNMENTS

RESULT 1
S18350
seed storage protein - barley
C/Species: Hordeum vulgare (barley)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C/Accession: S18350
R:Entwistle, J.; Knudsen, S.; Mueller, M.; Cameron-Mills, V.
Plant Mol. Biol. 17, 1217-1231, 1991
A/Title: Amber codon suppression: the in vivo and in vitro analysis of two C-hordein
A/Reference number: S18350; MUID:92032786; PMID:1932695
A/Accession: S18350
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-260 <EMT>
A/Cross-References: EMBL:X60037; NID:q19000; PIDN:CAA42642.1; PID:q19001
C/Superfamily: gliadin

Query Match
Best Local Similarity 91.7% Score 830.5; DB 2; Length 260;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 ROLNFCSELOSPQOSYLOQPPYLPQKPFYVQOPFPYQPFYELPQYQIP 61
DB 21 ROLNFCSELOSPQOSYLOQPPYLPQKPFYVQOPFPYQPFYELPQYQIP 80
QY 62 TPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 121
DB 81 TPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 140
QY 122 QITFQPPQOSYPPVQPPQPPQPPQPPQPPQ 150
DB 141 QITFQPPQOSYPPVQPPQPPQPPQPPQPPQ 168

RESULT 2
T06211
C-hordein precursor - barley
C/Species: Hordeum vulgare (barley)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C/Accession: T06211; A03355
R:Salnova, O.V.; Mekhedov, S.L.; Zheludin, L.G.; Khokhlova, T.A.; Anan'ev, E.V.
Genetika 29, 1070-1079, 1993
A/Title: Nucleotide sequence of the barley C-hordein gene.
A/Reference number: Z15537; MUID:93380629; PMID:8396543
A/Accession: T06211
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-310 <SAT>
A/Cross-References: EMBL:S6638; NID:q442523; PIDN:AAB28161.1; PID:q442524
R:Shewry, P.R.; Autran, J.C.; Nimmo, C.C.; Lew, E.J.L.; Kasarda, D.D.
Nature 286, 520-522, 1980
A/Reference number: A03228

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:57:50 ; Search time 150.754 Seconds
(without alignments)
4178.078 Million cell updates/sec

Title: US-09-743-533-18
Perfect score: 477
Sequence: 1 atgagagcaactaaaccttg.....gcataatgagatagtgctag 477

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	106.2	22.3	1635	US-09-864-761-20241	Sequence 20241, A
2	106.2	22.3	1973	US-09-864-761-3471	Sequence 3471, Ap
3	103	21.6	15720	US-10-025-380-1058	Sequence 1058, Ap
4	103	21.6	15720	US-09-922-217-1058	Sequence 1058, Ap
5	103	21.6	15720	US-09-833-263-1058	Sequence 1058, Ap
6	100.2	21.0	1075	US-09-864-761-19241	Sequence 19241, A
7	100.2	21.0	1403	US-09-864-761-2513	Sequence 2513, Ap
8	94.4	19.8	867	US-09-216-393-340	Sequence 340, App
9	94.4	19.8	867	US-09-216-393-342	Sequence 342, App
10	94.4	19.8	1397	US-09-216-393-343	Sequence 343, App
11	94.4	19.8	1397	US-09-216-393-345	Sequence 345, App
12	93.6	19.6	1236	US-10-077-584-3	Sequence 3, Appl1
13	92.6	19.4	446	US-09-864-761-20699	Sequence 20699, A
14	91.2	19.1	830	US-09-864-761-19531	Sequence 19531, A
15	77.4	16.2	479	US-09-864-761-768	Sequence 768, App
16	77.2	16.2	401	US-09-864-761-3936	Sequence 3936, App
17	77	16.1	155074	US-10-026-188-6	Sequence 6, Appl1
18	70.2	14.7	600	US-09-864-761-6690	Sequence 690, Ap
19	69.4	14.5	887	US-09-864-761-21484	Sequence 21484, A

C 20	69.4	14.5	1981	10	US-09-864-761-4745	Sequence 4745, Ap
C 21	67.8	14.2	2901	10	US-09-801-368-371	Sequence 371, App
C 22	67.4	14.1	520	9	US-10-184-644-332	Sequence 332, App
C 23	67.4	14.1	520	9	US-10-184-634-332	Sequence 332, App
C 24	66.6	14.0	493	10	US-09-864-761-5863	Sequence 5863, Ap
C 25	64.6	13.5	1322	9	US-09-789-054A-9	Sequence 9, Appl1
C 26	62.8	13.2	574	10	US-09-864-761-128	Sequence 128, App
C 27	62.8	13.2	669	10	US-09-864-761-17051	Sequence 17051, A
C 28	62.4	13.1	537	10	US-09-864-761-5592	Sequence 5592, App
C 29	61.8	13.0	350	10	US-09-864-761-23424	Sequence 23424, A
C 30	61.8	13.0	522	10	US-09-864-761-19900	Sequence 19900, A
C 31	60.8	12.7	549	9	US-10-255-536-208	Sequence 208, App
C 32	60.6	12.7	2656	10	US-09-842-552-82	Sequence 82, Appl
C 33	60.4	12.7	968	10	US-09-864-761-9629	Sequence 9629, Ap
C 34	60	12.6	925	9	US-09-735-056-1	Sequence 1, Appl1
C 35	59.2	12.4	470	10	US-09-864-761-3121	Sequence 3121, Ap
C 36	57.8	12.1	497	10	US-09-864-761-22361	Sequence 22361, A
C 37	57.8	12.1	1818	10	US-09-801-368-427	Sequence 427, App
C 38	57.4	12.0	1134	10	US-09-737-178-111	Sequence 111, App
C 39	57.4	12.0	1137	10	US-09-864-761-8390	Sequence 8390, App
C 40	57	11.9	537	10	US-09-864-761-8390	Sequence 8390, App
C 41	55.8	11.7	1089	10	US-09-925-302-301	Sequence 301, App
C 42	55.6	11.7	3150	9	US-09-938-842A-243	Sequence 243, App
C 43	55.6	11.7	4104	10	US-09-801-368-107	Sequence 107, App
C 44	55.4	11.6	476	10	US-09-864-761-6	Sequence 6, Appl1
C 45	55.2	11.6	768	9	US-09-938-842A-812	Sequence 812, App

ALIGNMENTS

RESULT 1
US-09-864-761-20241
: Sequence 20241, Application US/09864761
: Patent No. US2002048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Hanzel, David R.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecmeca-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263,6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20241
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-20241

```

```

Query Match      22.3%; Score 106.2; DB 10; Length 1635;
Best Local Similarity 53.2%; Pred. No. 1.1e-19;
Matches 225; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY 34 CAATACCAACAACATATATCTGCGGAGCATATGCAAAACCATATATACCGCAA 93
DB 501 CCATACCTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 560
QY 94 AAACATTTCCAGTGCAGCAACCGTTTCACACACCCCAACATATTTCCCTATCA 153
DB 561 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 620
QY 154 GAGAAATTTGTTCCCATATCAATACCAACCCCGACACCAACCAACCAACCAATTC 213
DB 621 CCACCATCTACTGCGGAGCATATGCAACCAACCAACCAACCAACCAACCAACCA 680
QY 214 CAACACCAACAACAACCTTCTCTGCGGAGCATATGCAAAACCATATATACCGCAA 273
DB 681 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 740
QY 274 CCATTTCCCGAGCCGAGCAACCAATTTCCCAACCAACCAATTTCCCTGCAACCA 333
DB 741 CTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 800
QY 334 CCATTTCCCGAGCCGAGCAACCAATTTCCCAACCAACCAATTTCCCTGCAACCA 393
DB 801 CCATACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 860
QY 394 TACCTGTGCAACCTCAACAGCATTTCTTAACACCTAACCAGTCCCAACAGCT 453
DB 861 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 920
QY 454 TCA 456
DB 921 TCA 923

```

RESULT 2

```

US-09-864-761-3471
; Sequence 3471, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

```

```

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3471
; LENGTH: 1973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471

```

```

Query Match      22.3%; Score 106.2; DB 10; Length 1973;
Best Local Similarity 53.2%; Pred. No. 1.2e-19;
Matches 225; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY 34 CAATACCAACAACATATATCTGCGGAGCATATGCAAAACCATATATACCGCAA 93
DB 655 CCATACCTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 714
QY 94 AAACATTTCCAGTGCAGCAACCGTTTCACACACCCCAACATATTTCCCTATCA 153
DB 715 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 774

```


QY	154	GAGGAATTTGTTTCCCGCAATATCAAAATACCAACCCCTACAGACCAAGCAACCAATTCGCC	213
Db	775	CCACCATCTACTGGCCGACGACGACGACCAACATCTACTACGACCAACGACCAACCATCTAG	834
QY	214	CAACCAACGACCAACCAACCTCTTCTCGGGGCCCAACCAACATTTCCGCTGGCAACGACCAACA	273
Db	835	CGACCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	894
QY	274	CCATTTTCCCGCAGCCGCCAAGAACAACCAATTTCCCGCAACGACATTTCCGCTGGCAACGACCAACA	333
Db	895	CTACACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	954
QY	334	CCATTTTCCCGCAGCCGCCAAGAACAACCAATTTCAAGCAATTAATTTTCAGCAACGCCCAACATCA	393
Db	955	CAATCACCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC	1014
QY	394	TACGCTGTGCAACGCTCAACGACGCAATTTCTTAACAACGCTCAACCAAGTCCGCCAACAAGCT	453
Db	1015	CCACGAC	1074
QY	454	TCA 456	
Db	1075	TCA 1077	

```

RESULT 3
US-10-025-380-1058
Sequence 1058, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tonglong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasser A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025.380
CURRENT FILING DATE: 2001-12-19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1058
LENGTH: 15720
TYPE: DNA
ORGANISM: Homo sapiens
US-10-025-380-1058

```

	Query Match	21.6%	Score 103	DB 9	length 15720
	Best Local Similarity	54.7%	Pred. No. 2,4e-18		
	Matches 252	Conservative 0	Mismatches 200	Indels 9	Gaps 2
Oy	7	CAACAAACCCCTTGAGCCAGAGAGTGGAAATCCACCAACAATATATCTGCGCAGGCA	66		
Db	4751	CCACCACCTCCAGCCCTCTCAACAAACACCACTCCGACGTCTCCAAACATACCAACAACA	4810		
Oy	67	TATCCACAAACCCATATCTACCCGAAAAACATTTCCAGTGCAGCAACGTTTACACA	126		
Db	4811	CCCTTCACAAACCAACCACTCCGACGCTCCGACACGACCAACCAACGACCCCTGCACAA	4870		
Oy	127	CCCCAACAAATTTTCCCCTATCTACAGAGGAATTTTCCCACATATCAATAACAAAC	186		

Db	4871	CCACCACTCCAGGCCCTGCACAGCACTACACCACTCCTCCACCAACCAAGCACTACACCC	4930
Qy	187	CCCGTACACCCACACAAACAT---TCGCCACACCAACACAACTCTTTCCTGGGCC	243
Db	4931	TTTCCACCAACACCACTCCAGCCCTCCACCAACCAACCAACCAACCCCTTCACCAACCA	4990
Qy	244	CAACAACCACTTCCCTCTGGCAACCAACCAACCAATTTGCCCA-----GCCCAAGAAACA	297
Db	4991	CCACTCTCCAGCCCTGCACCAACCAACCACTCCACCCCTCCATAACCAACCAACCAACCC	5050
Qy	298	ATTTCGCCAACACCACTTCCCTCTGGCAACCAACCAACCAATTTCCCAAGCCCAAGAACCA	357
Db	5051	CTCCACCAACCAACCACTCCAGCTCTCCAAATAACCAACCACTCCAGCCCTTCACCAACCA	5110
Qy	358	ATTCAACAATAATTTCCTCAGCAACCCCAACCAATCATTCCTGTGGCAACTCAACGCCA	417
Db	5111	CCATGACACCCCTTCACCAACCAACCAACCCCACTCTCCAAATAACCAACCAACCAACCC	5170
Qy	418	TTTCTCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	458
Db	5171	CTTCTCCACCAACCACTCCAGCCCTCCACCAACCAACCAACCAACCAACCAACCAACCA	5211

```

RESULT 4
US-09-922-217-1058
; Sequence 1058, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1058
; LENGTH: 15720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-217-1058

```

	Query Match	21.6%	Score 103	DB 10	Length 15720	
	Best Local Similarity	54.7%	Pred No. 2.4e-18			
	Matches 252	Conservative 0	Mismatches 200	Indels 9	Gaps 2	
QY	7	CAACTAAACCTTGGAGCCAGAGATTGGCAATTCACCACAACATCATCTATCTGCGGAGCA	66			
DB	4751	CCACCACGTCGAGCCCTCCACCAACACGACACTCCAGTCTCTCAACATTCACCAACACA	4810			
QY	67	TATCCACAAAACCCATATCTACGCCGAAAAACCATTTCCAGTGCACCAACCCGTTTCACCA	126			
DB	4811	CCCGCTCCACACGACGACCTCCGAGCCTCCCAACAGACGACACACGCCCTCCACCAA	4870			
QY	127	CCGCAACATATTTTCCCTATCTACAGAGCAATGTTTCCCATTAATCAATACCAAC	186			
DB	4871	CCAGCACTCCAGCCCTCCACAGACTACACCATCATCTCCACAAACGAGACTACGACC	4930			
QY	187	CCGCTACACCAACCAACCAT---TCCCCAACACGACAAACAACTCTTCTCGGGCC	243			
DB	4931	TTTCACGACGACGACCTCCGAGCCCTCCAGCAACGACGACGACGACCCCTCCACCAACA	4990			

OTHER INFORMATION: NT HIT: AL163201.2, EVALUATE 2.00e-19
US-09-864-761-19241

Query Match 21.0%, Score 100.2; DB 10; Length 1075;
Best Local Similarity 52.2%; Pred. No. 4.1e-18;
Matches 222; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 32 TGCATACCAACCAATCATATCTGCGGCGCATATCCCAAAACCAATATCTACCGG 91
DB 943 TACCACTACCAACCAATCATATCTGCGGCGCATATCCCAAAACCAATATCTACCGG 884
QY 92 AAAAACCATTTCAGTGAGCAACCGTTTCACACACCCCAACAAATATTTCCCTATCTAC 151
DB 883 CACACACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 824
QY 152 CAGAGGAATTTTTCCTCAATATATCAATACCAACCCCTTCAACCAACCAACCAATTC 211
DB 823 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 764
QY 212 CCGAACAACCAACCAACCACTCTTCTGCGGCGCAACCAATTCCTGCGGCAACCAAC 271
DB 763 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 704
QY 272 AACCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCAATTCCTGCGGCAACCAAC 331
DB 703 CATGACCAACCAACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 644
QY 332 AACCATTTCCCAAGCCCAAGAACCAATTTTCACCAACCAATTTTCACCAACCAACCAAT 391
DB 643 CACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACT 584
QY 392 CATACCTGTGCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 451
DB 583 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 524
QY 452 CTTC A 456
DB 523 CATCA 519

RESULT 7
US-09-864-761-2513/c
Sequence 2513, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2513
LENGTH: 1403
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL078472.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-2513

Query Match 21.0%, Score 100.2; DB 10; Length 1403;
Best Local Similarity 52.2%; Pred. No. 4.6e-18;
Matches 222; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 32 TGCATACCAACCAATCATATCTGCGGCGCATATCCCAAAACCAATATCTACCGG 91
DB 1293 TACCACTACCAACCAACCACTACCACTACCACTACCACTACCACTACCACTACCA 1234
QY 92 AAAAACCATTTCAGTGAGCAACCGTTTCACACACCCCAACAAATATTTCCCTATCTAC 151
DB 1233 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1174
QY 152 CAGAGGAATTTTCCCAATATCAATACCAACCCCTTCAACCAACCAACCAATTC 211
DB 1173 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1114
QY 212 CCAACAACCAACCAACCTCTTCTGCGGCGCAACCAACCAATTCCTGCGGCAACCAAC 271
DB 1113 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1054
QY 272 AACCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCAATTCCTGCGGCAACCAAC 331
DB 1053 CATGACCAACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 994
QY 332 AACCATTTCCCAAGCCCAAGAACCAATTTTCACCAACCAACCAATTTTCCTATCTAC 391
DB 993 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 934
QY 392 CATACCTGTGCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA 451
DB 933 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 874
QY 452 CTTC A 456
DB 873 CATCA 869

RESULT 8
US-09-216-393-340
; Sequence 340, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 340
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
US-09-216-393-340

Query Match 19.8%; Score 94.4; DB 10; Length 867;
Best Local Similarity 50.7%; Pred. No. 1.5e-16;
Matches 227; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 3 GAGGCACTAAACCTTGCAGCGCAAGATTGCATCACCACACATATATCTGCGCGCA 62
DB 388 GAGGAGGCTGCGTTTATTAGGCGCAAGAGGTCAAGAGGGAAGAAAAAGCTACACTACA 447
QY 63 GCCATATCCCAAAACCCATATCTACCGCAAAACCATTTCCAGTGCAGCAACGGTTTCA 122
DB 448 ACCAGTTTCATCCACAGTAGAGTAGACAGACGACGATCAACTACACTACTACACT 507
QY 123 CACACCCCAACATATTTTCCCTATCTACAGAGGAATTTGTTCCCAATATCAAAATCC 182
DB 508 ACCACCACTAGAGTAGACTAGCACTACGACACCAACACACAGCTACACACACACACA 567
QY 183 AACCCCTTACACACCAACACATTTCCGCCAACACACACACACTCTCTCTCGGCGC 242
DB 568 ACTACACCAACACACACACACACACACACACACACACTACACACACACACACACACAA 627
QY 243 CCACACACATTTCCCTGCGCAACACACACACATTTCCCGAGCCCAAGAACCAATTC 302
DB 628 ACAACACTAC 687
QY 303 CCACACACATTTCCCTGCGCAACACACACACATTTCCCGAGCCCAAGAACCAATTC 362
DB 688 ACAACACTACAGACTACAA 747
QY 363 ACAATATATTTTCCAGACACACACACACATATCTACCTGTGCACTTCAACGCAATTC 422
DB 748 CCACACACTAC 807
QY 423 TCACACACCTCAACCACTGTCGCCCAACAA 450
DB 808 ACAACTACACAGACTAGACACACATCTA 835

RESULT 9
US-09-216-393-342
; Sequence 342, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18

; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 342
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
US-09-216-393-342

Query Match 19.8%; Score 94.4; DB 10; Length 867;
Best Local Similarity 50.7%; Pred. No. 1.5e-16;
Matches 227; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 3 GAGGCACTAAACCTTGCAGCGCAAGATTGCATCACCACACATATATCTGCGCGCA 62
DB 388 GAGGAGGCTGCGTTTATTAGGCGCAAGAGGTCAAGAGGGAAGAAAAAGCTACACTACA 447
QY 63 GCCATATCCCAAAACCCATATCTACCGCAAAACCATTTCCAGTGCAGCAACGGTTTCA 122
DB 448 ACCAGTTTCATCCACAGTAGAGTAGACAGACGACGATCAACTACACTACTACACT 507
QY 123 CACACCCCAACATATTTTCCCTATCTACAGAGGAATTTGTTCCCAATATCAAAATCC 182
DB 508 ACCACCACTAGAGTAGACTAGCACTACGACACCAACACACAGCTACACACACACACA 567
QY 183 AACCCCTTACACACCAACACATTTCCGCCAACACACACACACTCTCTCTCGGCGC 242
DB 568 ACTACACCAACACACACACACACACACACACACACACTACACACACACACACACAA 627
QY 243 CCACACACATTTCCCTGCGCAACACACACACATTTCCCGAGCCCAAGAACCAATTC 302
DB 628 ACAACACTAC 687
QY 303 CCACACACATTTCCCTGCGCAACACACACACATTTCCCGAGCCCAAGAACCAATTC 362
DB 688 ACAACACTACAGACTACACACACACACACACACACACACACACACACACACACAA 747
QY 363 ACAATATATTTTCCAGACACACACACATATCTACCTGTGCACTTCAACGCAATTC 422
DB 748 CCACACACTAC 807
QY 423 TCACACACCTCAACCACTGTCGCCCAACAA 450
DB 808 ACAACTACACAGACTAGACACACATCTA 835

RESULT 10
US-09-216-393-343
; Sequence 343, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 343
; LENGTH: 1397
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (238)..(1104)
US-09-216-393-343

Query Match 19.8%; Score 94.4; DB 10; Length 1397;
Best Local Similarity 50.7%; Pred. No. 1.9e-16;

Matches 227: Conservative 0; Mismatches 221; Indels 0; Gaps 0;

```

OY 3 GAGGCAACTAACCCCTTGAGCCAGAGTTGCAATACACCAACATATATATGCGCGCA 62
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||  |||
DB 625 GAGGAGGCTGCGTTCTTAAAGAGAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 684
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 63 GGCATATCCCAAAACCCATATCTACCCGAAAACCATTTCCAGTGAACAACCGTTTCA 122
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 685 ACCAGTTTCATCCAAAGTACAGAGTACAGACGACGACGACGACGACGACGACGACG 744
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 123 CACAGCCCAACATATTTCCCTATCTACGAGAGATTTGTTCCCAATATCAATATCC 182
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 745 ACCGACCTAGAGTACTACTACACTACACCAACAACAACGACGACGACGACGACGAC 804
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 183 AACCCCTCAACAACCAACCAATTTCCCAACCAACCAACCAACCAACCAACCAACCA 242
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 805 ACTACACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 864
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 243 CCAGCAACCATTTCCCTGCGACGACGACGACGACGACGACGACGACGACGACGAC 302
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 865 ACAACCACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 924
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 303 CCAGCAACCATTTCCCTGCGACGACGACGACGACGACGACGACGACGACGACGAC 362
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 925 ACAACCACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 984
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 363 ACAATATATTTCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 422
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 985 CCAACCACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1044
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 423 TCAGCAACCTCAACCAACGATCCCAACAACA 450
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1045 ACAACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1072
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 11

```

US-09-216-393-345/C
; Sequence 345, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 345
; LENGTH: 1397
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
US-09-216-393-345

```

Query Match 19.8%; Score 94.4; DB 10; Length 1397;
 Best Local Similarity 50.7%; Pred. No. 1.9e-16;
 Matches 227: Conservative 0; Mismatches 221; Indels 0; Gaps 0;

```

OY 3 GAGGCAACTAACCCCTTGAGCCAGAGTTGCAATACACCAACATATATGCGCGCA 62
    |||||  |||  |||||  |||||  |||  |||  |||  |||  |||  |||  |||  |||
DB 773 GAGGAGGCTGCGTTCTTAAAGAGAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 714
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 63 GGCATATCCCAAAACCCATATCTACCCGAAAACCATTTCCAGTGAACAACCGTTTCA 122
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 713 ACCAGTTTCATCCAAAGTACAGAGTACAGACGACGACGACGACGACGACGACGACG 654
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 123 CACAGCCCAACATATTTCCCTATCTACGAGAGATTTGTTCCCAATATCAATATCC 182
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 653 ACCAGTTTCATCTACTACTACAGAGTACAGACGACGACGACGACGACGACGACGAC 594
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 183 AACCCCTCAACAACCAACCAATTTCCCAACCAACCAACCAACCAACCAACCAACCA 242
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

DB 593 ACTACACCAACAACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 534
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 243 CCAGCAACCATTTCCCTGCGACGACGACGACGACGACGACGACGACGACGACGAC 302
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 533 ACAACCACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 474
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 303 CCAGCAACCATTTCCCTGCGACGACGACGACGACGACGACGACGACGACGACGAC 362
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 473 ACAACCACTACGAGTACGACGACGACGACGACGACGACGACGACGACGACGAC 414
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 363 ACAATATATTTCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 422
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 413 CCAGCAACCATTTCCCTGCGACGACGACGACGACGACGACGACGACGACGACGAC 354
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 423 TCAGCAACCTCAACCAACGATCCCAACAACA 450
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 353 ACAACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 326
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 12

```

US-10-077-584-3
; Sequence 3, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
; APPLICANT: LINCOLN, SUSAN
; APPLICANT: KROBETSCH, SYLVIA
; APPLICANT: OUTIERO, TIAGO F.
; TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
; FILE REFERENCE: ARCD:36705
; CURRENT APPLICATION NUMBER: US/10/077,584
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269,157
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(513)
US-10-077-584-3

```

Query Match 19.6%; Score 93.6; DB 9; Length 1236;
 Best Local Similarity 51.2%; Pred. No. 3e-16;
 Matches 219: Conservative 0; Mismatches 209; Indels 0; Gaps 0;

```

OY 20 GCAGCCAGAGTTGCAATACCAACCAATATCTGCGCGACCAATATTCACAAACC 79
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 47 GCTTCCCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAAC 106
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 80 CATATATCCGCAAAACCATTTCCAGTGCAGCAACCGTTTCCAGCAACCCCAACATATT 139
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 107 AACAGAGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAG 166
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 140 TCCCTATCTACGAGAGATTTGTTCCCAATATCAATATCAACCCCTTACAGCAC 199
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 167 AGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAG 226
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 200 AACGACATTTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 259
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 227 AGCAACCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAACGAGCAAC 286
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 260 GGCAGCAACCAACCAACCAATTTCCAGCCCAAGAGCAACCAATTTCCCAACCAACCAAC 319
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 287 AACGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 346
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY 320 GGCAGCAACCAACCAACCAATTTCCCAACCAACCAACCAACCAACCAACCAACCAAC 379
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 347 AGCAACGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 406
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19531
; LENGTH: 830
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO 283851.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: SWISSPROT HIT: Q28614, EVALUE 2.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: Z45997.1, EVALUE 7.90e-01
US-09-864-761-19531
Query Match          19.1%; Score 91.2; DB 10: Length 830;
Best Local Similarity 51.2%; Pred. No. 11e-15;
Matches 213; Conservative 0; Mismatches 203; Indels 0; Gaps 0;
QY 34 CAATACCACCAACATCATATCTGCGCCAGCCCATATCCAGAAAACCATATCTACCGCAA 93
Db 579 CCATACACCAACATCATATCTGCGCCAGCCCATATCTACCGCAAACCATATCTACCGCAA 520
QY 94 AATCCATTTCCAGTGCAGCAACCGTTTCACACACCCCAACCAATATTTCCCTATCTACGA 153
Db 519 TCACCCACCAACATCATATCTGCGCCAGCCCATATCTACCGCAAACCATATCTACCGCAA 460
QY 154 GAGGAATTTGTTCCCAATATCAAAATACCAACCCCTTACGACACCAACCAATTTGCC 213
Db 459 CCACGACACCAACATCATATCTGCGCCAGCCCATATCTACCGCAAACCATATCTACCGCAA 400
QY 214 CAACACCAACCAACCTCTTCTCGGCGCCCAACCAATTTCCCTGCAACCAACAA 273
Db 399 CCACGACACCAACATCATATCTGCGCCAGCCCATATCTACCGCAAACCATATCTACCGCAA 340
QY 274 CCATTTCCGAGCCGCAAGAACCAATTTCCCAACCAACCATTTCCCTGCAACCAACAA 333
Db 339 CCACGACCAATTAATCTTCCACCAACCAACCAACCAACCAACCAACCAACCA 280
QY 334 CCATTTCCGAGCCGCAAGAACCAATTTCCCAACCAACCAATTTCCGAGCAACCCCAACAA 393
Db 279 CCACGACCAATTAATCTTCCACCAACCAACCAACCAACCAACCAACCAACCA 220
QY 394 TACCTGTGCAACTCAACGCAATTTCTCAACCAACCTCAACCAAGTCCGCCAACA 449
Db 219 CCACGACCAACCAATTAATCTTCCACCAACCAACCAACCAACCAACCAACCA 164
RESULT 15
```

```

US-09-864-761-768/c
; Sequence 768, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 768
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO 283851.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
US-09-864-761-768
Query Match          16.2%; Score 77.4; DB 10: Length 479;
Best Local Similarity 51.6%; Pred. No. 6e-12;
Matches 177; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:52:25 ; Search time 1662.16 Seconds

(without alignments)
4647.728 Million cell updates/sec

Title: US-09-743-533-18

Perfect score: 477
Sequence: 1 atgaaggaactaacacctg.....gcataatgagatgctag 477

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 805774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estlin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_iny:*
20: em_gss_pln:*
21: em_gss_vtl:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	321.6	67.4	567	14	BQ755360 EBbed07_SQ
2	319	66.9	585	10	BE195280 HVSMEH008
3	312.4	65.5	507	10	BE601687 HVSMEH009
4	287.4	60.3	936	12	BG365870 HVSME1000
5	246.8	51.7	515	9	AJ435523 AJ435523
6	246.2	51.6	420	9	AJ485191 AJ485191

C	7	245.8	51.5	360	9	AJ484592	AJ484592
C	8	236.8	49.6	300	9	AJ484591	AJ484591
	9	224.2	47.0	940	14	BQ609718	BQ609718
	10	219	45.9	595	14	BQ472381	BQ472381
	11	218.6	45.8	1008	13	B1947525	B1947525
	12	218	45.7	560	14	BQ245760	BQ245760
	13	217.8	45.7	611	14	BQ472392	BQ472392
	14	215	45.1	854	13	B1951055	B1951055
	15	213.8	44.8	706	14	BQ804665	BQ804665
	16	211	44.2	520	14	BQ472916	BQ472916
	17	197.8	41.5	494	12	BQ366682	BQ366682
	18	185.6	38.9	822	12	BQ366988	BQ366988
	19	182.6	38.3	762	14	BQ805800	BQ805800
	20	179	37.5	360	9	AJ479969	AJ479969
	21	179	37.5	420	9	AJ479968	AJ479968
	22	178.2	37.4	480	9	AJ479970	AJ479970
C	23	178.2	36.2	240	9	AJ481887	AJ481887
C	24	172.8	36.2	240	9	AJ484886	AJ484886
C	25	172.8	36.1	240	9	AJ484593	AJ484593
C	26	172	36.1	240	9	AJ484593	AJ484593
	27	168	35.2	360	9	AJ485127	AJ485127
	28	166.4	34.9	480	9	AJ479971	AJ479971
	29	162.4	34.0	516	14	BQ755868	BQ755868
	30	161.4	33.8	405	14	BQ755493	BQ755493
	31	161.4	33.8	480	9	AJ481886	AJ481886
	32	159.8	33.5	364	10	BE423915	BE423915
	33	158	33.1	742	14	BQ249043	BQ249043
	34	150.8	31.6	420	9	AJ485287	AJ485287
	35	150	31.4	606	14	BQ245034	BQ245034
	36	150	31.4	818	14	BQ806739	BQ806739
	37	148.4	31.1	497	12	BF293550	BF293550
	38	148.4	31.1	515	14	BQ606957	BQ606957
	39	148.4	31.1	518	10	BE402654	BE402654
	40	148.4	31.1	518	14	BQ608417	BQ608417
	41	148.4	31.1	574	14	BQ805851	BQ805851
	42	148.4	31.1	584	14	BQ244692	BQ244692
	43	148.4	31.1	592	14	BQ246427	BQ246427
	44	148.4	31.1	640	10	BE427016	BE427016
	45	148.4	31.1	640	10	BE427051	BE427051

ALIGNMENTS

RESULT 1
BQ755360 567 bp mRNA linear EST 26-JUL-2002
LOCUS EBbed07_SQ002.L09_R endospERM_28 DPA, no treatment, cv Optic,
DEFINITION EBbed07_Hordeum vulgare cDNA clone EBbed07_SQ002.L09 5', mRNA
Sequence.
ACCESSION BQ755360
VERSION BQ755360.1 GI:21963832
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.

REFERENCE

1 (bases 1 to 567)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.
Development of Barley Transcriptome Resources
Unpublished (2001)

TITLE

JOURNAL

COMMENT

Contact: Maugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: estescri.sari.ac.uk.
location/Qualifiers

FEATURES

1..567
/organism="Hordeum vulgare"

```

/cultivar="Optic"
/db.xref="taxon:4513"
/clone="EBed07.S0002.L09"
/clone_lib="endosperm, 28 DPA, no treatment, cv Optic, Ebed07"
/tissue_type="endosperm"
/dev_stage="28 DPA"
/lab_host="DH10B"
/Note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from endosperm tissue dissected from developing grains (28 days post anthesis) in glasshouse grown barley plants. Developed as part of the barley transcriptome resources of BBSRC/SFERAD funded cereal IGF (Investigating Gene Function) project."
Gene Function) project."

BASE COUNT      197 a      224 c      48 g      98 t
ORIGIN
Query Match      67.4%; Score 321.6; DB 14; Length 567;
Best Local Similarity 84.8%; Pred. No. 1e-59;
Matches 373; Conservative 0; Mismatches 64; Indels 3; Gaps 1;

OY 4 AGCAACTAAACCTTGACGCAAGATTGCATTCACACACATCATATCTGCCGAG 63
    |||||||
Db 106 AGGCACTAAACCTTGACGCAAGATTGCATTCACACACATCATATCTGCCGAG 165
    |||||||

OY 64 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGACACACCGTTTAC 123
    |||||||
Db 166 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGACACACCGTTTAC 225
    |||||||

OY 124 ACACCCCAACAAATTTTCCCTATCTACGAGGAATTTGCCAATTCAAATACCA 183
    |||||||
Db 226 ACACCCCAACAAATTTTCCCTATCTACGAGGAATTTGCCAATTCAAATACCA 285
    |||||||

OY 184 ACCCCCTACACCAACAACCATTTCCCAACACACACACACCTTTCTCGGCC 243
    |||||||
Db 286 ACCCCCTACACCAACAACCATTTCCCAACACACACACACCTTTCTCGGCC 345
    |||||||

OY 244 CAACACACATTTCCCTTGCAACCAACCATTTCCCAAGCCCAAGAACATTTCC 303
    |||||||
Db 346 CAACACACATTTCCCTTGCAACCAACCATTTCCCAAGCCCAAGAACATTTCC 405
    |||||||

OY 304 CAACACATTTCCCTTGCAACCAACCATTTCCCAAGCCCAAGAACATTTCC 363
    |||||||
Db 406 CAACACATTTCCCTTGCAACCAACCATTTCCCAAGCCCAAGAACATTTCC 465
    |||||||

OY 364 CAATTAATTTTTCAGCAACCAACCAATCATACCTGTGCAACCTTCT 423
    |||||||
Db 466 CAACCTTTTCTCAG--CCCAACCAACCATTTCCCTTGCAACCAACCAATTTCC 522
    |||||||

OY 424 CAACACCTCAACCAAGTCC 443
    |||||||
Db 523 CAGCCCAACCAACCAATTTCC 542
    |||||||

RESULT 2
BE195280      585 bp      mRNA      linear      EST 22-OCT-2001
LOCUS      HVSMEH0088M16f Hordeum vulgare 5-45 DAP spike EST library
DEFINITION      HVCDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0088M16f,
                mRNA sequence.
ACCESSION      BE195280
VERSION      BE195280
KEYWORDS      GI:16321166
SOURCE      EST
ORGANISM      Hordeum vulgare.
REFERENCE      Wing,R., Close,T.J., Kleinof,A., Wise,R., Begum,D., Fritsch,D., Yu
AUTHORS      Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
                ,R.D., Close,S.J., Oates,R. and Main,D.

```

```

TITLE      Development of a genetically and physically anchored EST resource
            for barley genomics: Morex 5-45 DAP spike cDNA library
JOURNAL      unpublished (2001)
COMMENT      On Jun 26, 2000 this sequence version replaced gi:13188066.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total bp bases = 341
            Seq primer: AATTACCCCTCACTAAAGG
            High quality sequence stop: 574.
            Location/Qualifiers
                1..585
                /organism="Hordeum vulgare"
                /cultivar="Morex"
                /db.xref="taxon:4513"
                /clone="HVSMEH0088M16f"
                /clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
                HVCDA0009 (5 to 45 DAP)
                /tissue_type="5-45 DAP Spike"
                /lab_host="SOLR"
                /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give plasmid SK(-) cDNA phagemids (Choi) in the TJ Close lab at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Fritsch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Fritsch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see close TJ, Wing R, Kleinof A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT      208 a      228 c      47 g      102 t
ORIGIN
Query Match      66.9%; Score 319; DB 10; Length 585;
Best Local Similarity 84.8%; Pred. No. 3.7e-59;
Matches 363; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

OY 4 AGCAACTAAACCTTGACGCAAGATTGCATTCACACATCATATCTGCCGAG 63
    |||||||
Db 118 AGGCACTAAACCTTGACGCAAGATTGCATTCACACATCATATCTGCCGAG 177
    |||||||

OY 64 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGACACACCGTTTAC 123
    |||||||
Db 178 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGACACACCGTTTAC 237
    |||||||

OY 124 ACACCCCAACAAATTTTCCCTATCTACGAGGAATTTGCCAATATATACCA 183
    |||||||
Db 238 ACACCCCAACAAATTTTCCCTATCTACGAGGAATTTGCCAATATATACCA 297
    |||||||

OY 184 ACCCCCTACACCAACCAACCATTTCCCAAGCCCAAGAACATTTCTCGGCC 243
    |||||||
Db 298 ACCCCCTACACCAACCAACCATTTCCCAAGCCCAAGAACATTTCTCGGCC 357
    |||||||

OY 244 CAACACATTTCCCTTGCAACCAACCATTTCCCAAGCCCAAGAACATTTCC 303
    |||||||

```


ACCESSION AJ485191 GI:21201146
 VERSION AJ485191.1
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 420)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 TITLE Barley EST's
 JOURNAL Unpublished (2002)
 COMMENT Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O.Box 56 (Vikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES
 source Location/Qualifiers
 1..420
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 /clone="S0001100164F01F1"
 /dev_stage="Developing seed"
 /note="12,15,18 days after pollination"

BASE COUNT 145 a 156 c 43 g 76 t

ORIGIN

Query Match 51.6%; Score 246.2; DB 9; Length 420;
 Best Local Similarity 87.4%; Pred. No. 1.9e-43;
 Matches 291; Conservative 0; Mismatches 18; Indels 24; Gaps 1;

QY 4 AGGCACTAAACCTTGACGACCAAGAGTGTGATCAGCAACCAATCATATCTCCGCG 63
 DB 111 AGGCACTAAACCTTGACGACCAAGAGTGTGATCAGCAACCAATCATATCTCCGCG 170
 DB 171 CCATATCCCAAAACCCATATCTACCGCAACCAATTTCCAGTGACCAACCGTTTCC 230
 QY 64 CCATATCCCAAAACCCATATCTACCGCAAAACCAATTTCCAGTGACCAACCGTTTCC 123
 DB 171 CCATATCCCAAAACCCATATCTACCGCAACCAATTTCCAGTGACCAACCGTTTCC 230
 QY 124 ACACCCCAACCAATTTCCCTGATCTACGAGAGATTTTCCCATATCAATATGCA 183
 DB 231 ACACCCCAACCAATTTCCCTGATCTACGAGAGATTTTCCCATATCAATATGCA 290
 QY 184 ACCCCCTTACACCAACCAATTTCCCGCAACCAACCAACCTTTCTCTGGGCC 243
 DB 291 ACCCCCTT-----ACAACCAACCAACCTTTCTCTGGGCC 326
 QY 244 CAACCAACCAATTTCCCGTGGCAACCAACCAATTTCCCGCGCCAGAACCAATTTCC 303
 DB 327 CAACCAACCAATTTCCCGTGGCAACCAACCAATTTCCCGCGCCAGAACCAATTTCC 386
 QY 304 CAACCAACCAATTTCCCGTGGCAACCAACCAACCAACCAACCAACCAACCAACCA 336
 DB 387 CAGCAACCAACCAATTTCCCGTGGCAACCAACCAACCAACCAACCAACCAACCAACCA 419

RESULT 7
 AJ484592 360 bp mRNA linear EST 24-MAY-2002
 LOCUS AJ484592 S00011 Hordeum vulgare cDNA clone S0001100047B07F1, mRNA
 DEFINITION sequence.
 ACCESSION AJ484592
 VERSION AJ484592.1 GI:21200548
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 360)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 TITLE Barley EST's

JOURNAL Unpublished (2002)
 COMMENT Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O.Box 56 (Vikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES
 source Location/Qualifiers
 1..360
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 /clone="S0001100047B07F1"
 /dev_stage="Developing seed"
 /note="12,15,18 days after pollination"

BASE COUNT 67 a 37 c 136 g 120 t

ORIGIN

Query Match 51.5%; Score 245.8; DB 9; Length 360;
 Best Local Similarity 87.6%; Pred. No. 2.2e-43;
 Matches 290; Conservative 0; Mismatches 17; Indels 24; Gaps 1;

QY 4 AGGCACTAAACCTTGACGACCAAGAGTGTGATCAGCAACCAATCATATCTCCGCG 63
 DB 309 AGGCACTAAACCTTGACGACCAAGAGTGTGATCAGCAACCAATCATATCTCCGCG 250
 QY 64 CCATATCCCAAAACCCATATCTACCGCAAAACCAATTTCCAGTGACCAACCGTTTCC 123
 DB 249 CCATATCCCAAAACCCATATCTACCGCAACCAATTTCCAGTGACCAACCGTTTCC 190
 QY 124 ACACCCCAACCAATTTCCCTGATCTACGAGAGATTTTCCCATATCAATATGCA 183
 DB 189 ACACCCCAACCAATTTCCCTGATCTACGAGAGATTTTCCCATATCAATATGCA 130
 QY 184 ACCCCCTTACACCAACCAATTTCCCGCAACCAACCAACCTTTCTCTGGGCC 243
 DB 129 ACCCCCTT-----ACAACCAACCAACCTTTCTCTGGGCC 94
 QY 244 CAACCAACCAATTTCCCGTGGCAACCAACCAATTTCCCGCGCCAGAACCAATTTCC 303
 DB 93 CAACCAACCAATTTCCCGTGGCAACCAACCAATTTCCCGCGCCAGAACCAATTTCC 34
 QY 304 CAACCAACCAATTTCCCGTGGCAACCAACCAACCAACCAACCAACCAACCAACCA 334
 DB 33 CAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3

RESULT 8
 AJ484591 300 bp mRNA linear EST 24-MAY-2002
 LOCUS AJ484591 S00011 Hordeum vulgare cDNA clone S0001100043B09F1, mRNA
 DEFINITION sequence.
 ACCESSION AJ484591
 VERSION AJ484591.1 GI:21200547
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 300)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 TITLE Barley EST's
 JOURNAL Unpublished (2002)
 COMMENT Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O.Box 56 (Vikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES
 source Location/Qualifiers
 1..300
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 /clone="S0001100043B09F1"

BASE COUNT 49 a 26 c 116 g 109 t
 ORIGIN /clone.lib="S00011"
 /dev_stage="Developing seed"
 /note="12,15,18 days after pollination"

Query Match 49.6%; Score 236.8; DB 9; Length 300;
 Best Local Similarity 87.3%; Pred. No. 1.8e-41;
 Matches 281; Conservative 0; Mismatches 17; Indels 24; Gaps 1;

QY 13 AACCTTGACGCAAGAGTTGCAATCAGCAACAATCATATCGCCGACGCAATATCA 72
 Db 300 AACCTTGACGCAAGAGTTGCAATCAGCAACAATCATATCGAGGAGGCAATATCA 241
 QY 73 CAAACCCATATACGCAAGCAAAACCATTTCCAGTACAGCAACGTTTACAGACCCCA 132
 Db 240 CAAACCCATATACGCAAGCAAAACCATTTCCAGTACAGCAACGTTTACAGACCCCA 181
 QY 133 CAATTTTCCCTATTCACAGAGAAATTTTCCCATATTCATACCAACCCCTA 192
 Db 180 CAATTTTCCCTATTCACAGAGAAATTTTCCCATATTCATACCAACCCCTA 122
 QY 193 CAACCAACAACCAATTTCCCAACAACAACCTTCTCTGCGCCCAACAACCA 252
 Db 121 -----ACAACCAACAACCAATTTCTCTGCGCCCAACAACCA 85
 QY 253 TTCCCTTGACCAACAACAACCAATTTTCCCAACAACAACCAATTTCCCAACAACCA 312
 Db 84 TTCCCTTGACCAACAACAACCAATTTTCCCAACAACAACCAATTTCCCAACAACCA 25
 QY 313 TTCCCTTGACCAACAACAACCAATTTTCCCAACAACAACCAATTTCCCAACAACCA 334
 Db 24 CAACCAACAACCAACAACAACCAATTTTCCCAACAACAACCAATTTCCCAACAACCA 3

RESULT 9
 LOCUS BQ609718 940 bp mRNA linear EST 25-JUN-2002
 DEFINITION BRY_5804 wheat EST endosperm library Triticum aestivum cDNA 5',
 mRNA sequence.

ACCESSION BQ609718
 VERSION BQ609718.1 GI:21559057
 KEYWORDS EST
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.
 1 (bases 1 to 940)
 Clarke,B., Lambrecht,M. and Rhee,S.
 Assessing the utility of Arabidopsis genomic information for
 interpreting wheat EST sequences
 Unpublished (2002)

JOURNAL The Arabidopsis Information Resource
 CONTACT: Lambrecht M
 COMMENT Carregie Institution of Washington, Dept. of Plant Biology
 260 Panama Street, Stanford, CA 94305, USA
 Tel: 1 650 325 1521 x 251
 Fax: 1 650 325 3748
 Email: rhee@coma.stanford.edu.

FEATURES
 source Location/Qualifiers
 1..940
 /organism="Triticum aestivum"
 /cultivar="WYuna"
 /db_xref="taxon:4565"
 /clone_lib="wheat EST endosperm library"
 /tissue_type="endosperm"
 /dev_stage="developing endosperm tissue 8, 10 and 12 DPA
 (days post anthesis)"
 /note="Vector: Bluescript II SK(-)"

BASE COUNT 340 a 365 c 90 g 145 t

Query Match 47.0%; Score 224.2; DB 14; Length 940;
 Best Local Similarity 73.5%; Pred. No. 1.3e-38;
 Matches 324; Conservative 0; Mismatches 93; Indels 24; Gaps 2;

QY 4 AGCAACTAAACCCCTGAGCAGAGTTGCAATCAGCAACAATCATATCTGCGGAG 63
 Db 166 AGCAACTAAACCCCTGAGCAGAGTTGCAATCAGCAACAATCATATCTGCGGAG 210
 QY 64 CCATTTCCCAAAACCCATATTCACGCAAGAAACCAATTTCCAGTACGCAACGTTTAC 123
 Db 211 CCGTTTCCCAAAACCCATATTCACGCAAGAAACCAATTTCCAGTACGCAACGTTTAC 270
 QY 124 ACACCCCAACAATTTTCCCTATTCACGCAAGAAATTTTCCCAATATCAATATCA 183
 Db 271 ACACCCCAACAATTTTCCCTATTCACGCAAGAAATTTTCCCAATATCAATATCA 330
 QY 184 ACCCCCTTACACCAACAACAACCAATTTTCCCAACAACAACCAATTTTCCCAACA 243
 Db 331 ACCCCCTTACACCAACAACAACCAATTTTCCCAACAACAACCAATTTTCCCAACA 390
 QY 244 CAACAACCAATTTTCCCTGCAACAACAACAATTTTCCCAACAACAATTTTCCCAACA 303
 Db 391 CAACAACCAATTTTCCCTGCAACAACAACAATTTTCCCAACAACAATTTTCCCAACA 450
 QY 304 CAACAACA-----TTCCCTGCAACAACAACAATTTTCCCAACAACAATTTTCCCAACA 354
 Db 451 CAACAACAACAATTTTCCCTGCAACAACAACAATTTTCCCAACAACAATTTTCCCAACA 510
 QY 355 CCATTTCAACAATATTTTCCCAACAACAACAATTTTCCCAACAACAATTTTCCCAACA 414
 Db 511 CAACAACAACAATATTTTCCCAACAACAACAATTTTCCCAACAACAATTTTCCCAACA 570
 QY 415 CCATTTCCCAACAACAATTTTCCCAACAACAATTTTCCCAACAACAATTTTCCCAACA 435
 Db 571 CCATTTCCCAACAACAATTTTCCCAACAACAATTTTCCCAACAACAATTTTCCCAACA 591

RESULT 10
 LOCUS BQ472381 595 bp mRNA linear EST 31-MAY-2002
 DEFINITION HB09M16T BC Hordeum vulgare cDNA clone HB09M16 5-PRIME, mRNA
 sequence.

ACCESSION BQ472381
 VERSION BQ472381.1 GI:21284510
 KEYWORDS EST
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 595)
 Radchuk,V., Zhang,H., Weschke,M., Potokina,E. and Wobus,U.
 Barley ESTs from developing seeds
 Unpublished (2002)

JOURNAL Molecular Markers Group, Department Genbank
 CONTACT: Stein Nils
 COMMENT Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3 06466, Gatersleben, Germany
 Tel: 039462-5522
 Fax: 039462-5595
 Email: stein@ipk-gatersleben.de
 Insert length: 595 Std Error: 0.00
 Plate: 9 row: M column: 16
 Seq primer: 73.

FEATURES
 source Location/Qualifiers
 1..595
 /organism="Hordeum vulgare"
 /cultivar="Baiké"
 /db_xref="taxon:4513"
 /clone_lib="HB09M16"
 /tissue_type="developing caryopsis"
 /dev_stage="8-15 DAP (days after pollination)"

BASE COUNT 340 a 365 c 90 g 145 t

Db 409 CATTCCCTAGCAACCAACCAATATTTCCCAACCAACCAACCAATTCCTCCCTG 468
 Oy 347 CCCAGAACCAATTCACCAATATTTCCAGCAACCAACCAATTCCTCCCTG 406
 Db 469 AACCCCAACAA---CCACATCAACCTTTCTCGGCCCCCAACCAACCAATTCCTCCCTG 525
 Oy 407 CTCACAGCAATTT 420
 Db 526 CACACCAACCAATTT 539

RESULT 12
 B0245760 560 bp mRNA linear EST 03-MAY-2002
 LOCUS TAE15020E06R TAE15 Triticum aestivum cDNA clone TAE15020E06R, mRNA
 DEFINITION
 ACCESSION B0245760
 VERSION B0245760.1 GI:20441636
 KEYWORDS EST
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Triticum.
 1 (bases 1 to 560)
 Cloutier S.
 Wheat functional genomics - glenlea developing seeds cDNA libraries
 Unpublished (2002)
 CONTACT: Dr. Sylvie Cloutier
 Cereal Research Centre, Agriculture and Agri-food Canada
 195 Dufferin Rd, Winnipeg, MB, Canada R3T 2M9
 Tel: (204) 983-2340
 Fax: (204) 983-4604
 Email: scloutier@agr.ca
 was cloned directionally, not all sequences generated with reverse
 primer were from the 5' end (same with forward primer and 3' end).
 Average insert size is >1.4 kb
 Plate: 020 row: E column: 06
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..560
 /organism="Triticum aestivum"
 /cultivar="Glenlea"
 /db_xref="taxon:4565"
 /clone="TAE15020E06R"
 /clone_1ib="TAE15"
 /tissue_type="developing seeds"
 /dev_stage="15 days after anthesis"
 /lab_host="E. coli DH10B"
 /note="Vector: PCMV-SPORT6.0 (Invitrogen Technologies);
 Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
 of cultivar Glenlea 15 days post-anthesis"

BASE COUNT 192 a 240 c 41 g 87 t
 ORIGIN

Query Match 45.7% Score 218; DB 14; Length 560;
 Best Local Similarity 74.1%; Pred. No. 2,5e-37;
 Matches 329; Conservative 0; Mismatches 85; Indels 30; Gaps 3;

Oy 25 CAGAGTTCATCAACCAACCAATATTCGCGGCGCATATTCGCAAAACCAATAT 84
 Db 68 CATATCTTTTCTATCAACCAACCAATTCGAGGAGCATATTCGCAAAACCAATAT 127
 Oy 85 CAGAGCAAAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 144
 Db 128 TATCTACAGCAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 187
 Oy 145 TATCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 204
 Db 188 GAGCATTCACCAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 247
 Oy 205 CCAATTC-----CCCAACCAACCAACCAACCAACCAACCAACCAACCAATTC 255

Db 248 CCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAATTT 307
 Oy 256 CCGTGGCAACCAACCAACCAATTCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAG 306
 Db 308 CCGTGGCAACCAACCAACCAATTCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAG 367
 Oy 307 CAACCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAATTC- 361
 Db 368 CAGCATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAATTCCTCCAG 427
 Oy 362 -----AACCAATATTTTCAGCAACCAACCAACCAACCAACCAACCAACCAACG 414
 Db 428 CATCAAGCAACCAATATTCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 487
 Oy 415 CCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAATTCCTCCAG 438
 Db 488 CCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAATTCCTCCAG 511

RESULT 13
 B0472392 611 bp mRNA linear EST 31-MAY-2002
 LOCUS HB09M02T BC Hordeum vulgare cDNA clone HB09M02.5-PRIME, mRNA
 DEFINITION
 ACCESSION B0472392
 VERSION B0472392.1 GI:21284521
 KEYWORDS EST
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 611)
 Radchuk V., Zhang H., Wesche W., Potokina E. and Wobus U.
 Barley ESTs from developing seeds
 Unpublished (2002)
 CONTACT: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 611 Std Error: 0.00
 Plate: 9 row: M column: 2
 Seq primer: T3.
 Location/Qualifiers
 1..611
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HB09M02"
 /clone_1ib="BC"
 /tissue_type="developing caryopsis"
 /dev_stage="8-15 DAP (days after pollination)"
 /lab_host="X110-Gold"
 /note="Vector: pBluescript SK-; Site_1: EcoRI (5'-end of
 cDNA); Site_2: XhoI (3'-end of cDNA); developing caryopsis
 , 8-15 DAP(days after pollination) due to a cloning
 artefact caused by the kit. In most cases the EcoRI site
 is NOT present, as well as the EcoRI adapter used for
 cloning. To excise the insert, restriction sites upstream
 EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
 due to the cloning system used blue/white selection for
 recombinants is not 100% reliable."

BASE COUNT 198 a 244 c 57 g 112 t
 ORIGIN

Query Match 45.7% Score 217.8; DB 14; Length 611;
 Best Local Similarity 71.3%; Pred. No. 2,9e-37;
 Matches 323; Conservative 0; Mismatches 112; Indels 18; Gaps 2;

ACCESSION	BQ804665	GI:2202850
VERSION	BQ804665.1	
KEYWORDS	EST.	
SOURCE	bread wheat.	
ORGANISM	Triticum aestivum	

REFERENCE

AUTHORS

TITL E

JOURNAL
COMMENT

FEATURES
SOU

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	SOURCE	
BO0804665	BO0804665.1	GI:22028850	EST.	Bread wheat.	1 (bases 1 to 706)	Altshbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronlin,K., Crossman,C., Fenton,R.D., Iazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.	The structure and function of the expressed portion of the wheat genomes - Developing grains cDNA library	Unpublished (2002)	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105955773 Fax: 5105958818 Email: oanderson@pw.usda.gov	Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: 5K primer.	Location/Qualifiers 1. 706 /organism="Triticum aestivum" /cultivar="Butte 86" /db_xref="taxon:4565" /clone="WHB357_D02_H03" /clone_lib="Wheat developing grains cDNA library" /tissue_type="whole grains" /dev_stage="3-44 days post anthesis seed" /lab_host="E. coli SOLR" /note="Vector: lambda ZAP II, excised phagemid: Site 1: EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 240C/170C day/night, well-watered, with post-anthesis fertilizer, Environment 2) 240C/170C day/night, well-watered, without post-anthesis fertilizer, Environment 3) 370C/170C day/night, well-watered, with post-anthesis fertilizer, Environment 4) 370C/170C day/night, well-watered, without post-anthesis fertilizer, Environment 5) 370C/170C day/night plus drought, with post-anthesis fertilizer, Environment 6) 370C/170C day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altshbach and K. Cronlin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give plasmid SK(-) phagemids in the TJ Close lab (Chin, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."	246 a 325 c 39 g 96 t

Query Match	44.8%;	Score 213.8;	DB 14;	Length 706;
Best Local Similarity	74.9%;	Pred. No. 2.2e-36;		
Matches 320; Conservative	0;	Mismatches 77;	Indels 30;	Gaps 3;

42 ACACATCATATCTGCGGAGCCATATCCACAAAACCCATATCTAACCGCAAAAACCATTT 101

Db	2	ACAAACAACCAATTCCACAGCAACCATATCCACAAACAATATCATCATCAGCAAGCAACCATATA	61
Qy	102	TCGAGTGGAGCAACCGTTTCACACAGCCGCAACCAATATTTCCCTATTTACAGAGGAAATT	161
Db	62	TCCATTTCAACAACCAATTTCACACCCCAACCAACAATTTTCCCGAGCAATCAACAACCC	121
Qy	162	GTTTTCGCCAATATCAAAATACCAACCCGCGCTAGACAGCAACAACCAATTC-----CC	212
Db	122	ATTTATCCGAGCGCCCAACAACGAGCGCCGCTATCAACGCAACAACAACCATTTCCCGACGCAAC	181
Qy	213	CCAAACAACCAACAACCTCTTCTCGGCGCCCAACAACAATTCCTCGGCAACCAACACA	272
Db	182	CCAAACAACCAACAACCTTTTCCAGACCGCAACAACAATTTCCCTGGCAACCAACACA	241
Qy	273	ACCAATTTTCCGACCGCCCAAGAACAAT-----TCCGCAACAACAATTTCCCTGGCA	323
Db	242	ACCAATTTTCCGACCGCAACAATCTTTCCCTGCAACAACAACAATTTTCCCGCGCA	301
Qy	324	ACCAACAACAACATTTTCCCGACCGCCCAAGAACAATTC-----AACAAATTAAT	371
Db	302	ACCCCAACAACAATTTCCCGACCGCCCAACTACAATTTCCCGCAACAACAACAACAATTAAT	361
Qy	372	TTTCCGAGAACCCCAACAATCATACCTGTGCAACCTCAAGAGCAATTTCTCAACAAC	431
Db	362	TCCCGCAGCAACCCCAACAACCAATTTCTCTGATTCGCAACAACAATTTCCCGCGCAAC	421
Qy	432	TCAACA 438	
Db	422	CCAAACA 428	

Search completed: May 25, 2003, 20:17:14
Job time : 1666.16 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:35:40 : Search time 48 Seconds

(without alignments)
438.617 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906
Sequence: 1 MRQLNPGSELQSPQGSYLQ.....PFPQPPVPVPOQASCIWSMV 158

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	906	100.0	158	21	AAV54568 A synthetically tr
2	262.5	29.0	1162	21	AAV56255 Kaposi's sarcoma-a
3	262.5	29.0	1162	21	AAV58500 HHV8 ORF 73 protei
4	262.5	29.0	1162	22	AAAB6231 Amino acid sequenc
5	262.5	29.0	1162	23	AAAB6262 Kaposi's sarcoma-a
6	255.5	28.2	369	19	AAAB6267 Mature durum wheat
7	244.5	27.0	1362	22	AAU38416 Salmorella typhi c
8	234	25.8	266	22	AAU01799 Wheat A-glialdin.
9	217	24.0	1329	22	AAU34519 E. coli cellular p
10	213	23.5	308	21	AAAB4459 Plant viral moveime

11	204	22.5	1299	22	AAW24322 Human EST encoded
12	204	22.5	1404	13	AAW26049 MSF precursor. Sy
13	204	22.5	1404	22	AAAB60568 Human megakaryocyt
14	204	22.5	1404	22	AAAB29773 Human megakaryocyt
15	204	22.5	1415	22	AAU32262 Novel human secret
16	202.5	22.4	717	22	ABG17438 Novel human diagno
17	200.5	22.1	647	22	ABAB7093 Drosophila melanog
18	200.5	22.1	672	22	ABBS9473 Drosophila melanog
19	196.5	21.7	2285	22	ABBS3057 Drosophila melanog
20	196	21.6	757	22	AAAG85017 Shrimp white spot
21	195	21.5	446	22	ABAB70663 Drosophila melanog
22	194	21.4	882	22	ABBS1046 Drosophila melanog
23	190	21.0	1798	22	ABAB71695 Drosophila melanog
24	187	20.6	526	21	AAAB12717 Streptococcus pneu
25	187	20.6	987	21	AAAB33064 Human tyrosine pho
26	187	20.6	1253	21	AAAB29663 Human tyrosine pho
27	187	20.6	1264	21	AAAB29664 Human tyrosine pho
28	187	20.6	1274	20	AAAB9253 Human Alp. Homo s
29	187	20.6	1583	21	AAAB29662 Human tyrosine pho
30	187	20.6	1636	21	AAAB29661 Human histidine do
31	184.5	20.4	450	22	ABAB71041 Drosophila melanog
32	184	20.3	667	19	AAAB48760 BOP1 protein. Mus
33	183.5	20.3	371	19	AAAB64535 Human leukocyte ce
34	183.5	20.3	371	23	ABAB61444 Human NF-kB activa
35	182.5	20.1	522	21	AAAG13630 Arabidopsis thalia
36	182.5	20.1	539	21	AAAG13629 Arabidopsis thalia
37	182.5	20.1	540	21	AAAG13628 Arabidopsis thalia
38	180	19.9	148	21	AAI92720 C. albicans hyphal
39	179.5	19.8	961	22	ABAB7326 Drosophila melanog
40	177.5	19.6	180	21	AAAB45073 Rat amelogenin pro
41	175	19.3	194	33	AAAY45072 Human polypeptide
42	175	19.3	748	22	AAAM40299 Human polypeptide
43	175	19.3	780	22	AAAM42065 Human polypeptide
44	174	19.2	439	13	AAAB28150 Sugar beet chitina
45	173.5	19.2	688	22	ABAB67404 Drosophila melanog

ALIGNMENTS

RESULT 1	AAV54568	AAV54568 standard. Protein; 158 AA.
ID	AAV54568	
XX	AAV54568	
AC	AAV54568	
XX	AAV54568	
DT	25-APR-2000 (first entry)	
XX		
DE	A synthetically truncated C hordein protein (glutenin).	
XX		
KW	C hordein gene; glutenin; seed storage protein; gluten; bread; pasta;	
KW	noodel; breakfast cereal; snack food; cake; pastry; flour based sauce;	
KW	film; coating; adhesive; building material; packaging material; grain.	
XX		
OS	Hordeum vulgare.	
XX	Synthetic.	
FT	Key	Location/Qualifiers
FT	Misc-difference 20	/note= "encoded by CCG"
FT	Misc-difference 105	/note= "encoded by TTC"
FT	Misc-difference 106	/note= "encoded by CCC"
FT	Misc-difference 107	/note= "encoded by TGG"
FT	Misc-difference 108	/note= "encoded by CAA"
FT	Misc-difference 113	/note= "encoded by TTT"
FT	Misc-difference 114	/note= "encoded by CCC"
FT	Misc-difference 115	/note= "encoded by CCC"

CC with the latency cycle of this virus. Potential antiviral treatments for
CC the above mentioned diseases may therefore be based on LANA deregulation.

SO Sequence 1162 AA:

Query Match 29.0%; Score 262.5; DB 21; Length 1162;

Best Local Similarity 46.8%; Pred. No. 5.9e-15;

Matches 74; Conservative 10; Mismatches 51; Indels 23; Gaps 8;

OY 2 ROUNPCSGE-LQSPQSYLQGPYQNPYLQKPFPPVQDPFTTPOQYFPLPEEL-FPQYQ 59
DB 453 QQQQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ 502
OY 60 IFTPLQPPQPPQOQ--PQOPLPRPQOQPFPPQOQEPQOQEPQOQEPQOQEPQOQ 115
DB 503 EPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ 562
OY 116 FPQO--PQOQIFQOQPOQSYPVQ--PQOQFPQOQPPQVQ 149
DB 563 EPQOQEPQOQ---REFQOQREPOQREPOQREPOQREPOQ 597

RESULT 3

AAV58500
ID AAV58500 standard; Protein: 1162 AA.

XX AAV58500;

DT 10-APR-2000 (first entry)

DE HHV8 ORF 73 protein, SEO ID NO:21.

KW HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen;

KM antigen.

OS Human herpesvirus type 8.

FT Key Location/Qualifiers

FT Misc-difference 96 /Label= unknown

PN WO9961909-A2.

PD 02-DEC-1999.

XX 26-MAY-1999; 99WO-US11407.

PR 26-MAY-1998; 98US-0086695.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Pau CP;

DR WPI: 2000-097142/08.

PS New methods and compositions for the detection of human herpesvirus

XX Claim 2: Page 59-62; 68pp; English.

CC Sequences AAV58480-Y5832 represent immunogenic polypeptides derived
CC from human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC invention relates to a novel method of detecting the presence of human
CC herpesvirus 8 in a biological sample using peptides representative of
CC dominant antigenic regions of HHV8. The method comprises contacting one
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing
CC biological sample, and detecting the formation of a complex between the
CC peptide and the antibody. The presence of a peptide-antibody complex
CC indicates the presence of human herpesvirus 8. The detection of HHV8
CC infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC HHV8-specific antibodies are useful therapeutically when for the passive
CC immunisation of a human against HHV8 infection, thereby reducing HHV8
CC related disease. The detection assays are highly specific, sensitive and

CC accurate. Early detection and treatment of Kaposi's sarcoma could
CC diminish the severity of symptoms related to AIDS and the sensitive
CC techniques could reduce erroneous characterisations of skin disorders.
CC Previous assays for HHV8 antibodies such as immunofluorescence assays,
CC immunoblots and enzyme immunoassays lack the sensitivity and accuracy
CC needed for reliable diagnosis of Kaposi's sarcoma. Further advantages
CC of the assays are that reproducible results are obtained and the method
CC is suitable for rapid throughput and screening of samples economically.

SO Sequence 1162 AA:

Query Match 29.0%; Score 262.5; DB 21; Length 1162;

Best Local Similarity 46.8%; Pred. No. 5.9e-15;

Matches 74; Conservative 10; Mismatches 51; Indels 23; Gaps 8;

OY 2 ROUNPCSGE-LQSPQSYLQGPYQNPYLQKPFPPVQDPFTTPOQYFPLPEEL-FPQYQ 59
DB 453 QQQQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ 502
OY 60 IFTPLQPPQPPQOQ--PQOPLPRPQOQPFPPQOQEPQOQEPQOQEPQOQEPQOQ 115
DB 503 EPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ 562
OY 116 FPQO--PQOQIFQOQPOQSYPVQ--PQOQFPQOQPPQVQ 149
DB 563 EPQOQEPQOQ---REFQOQREPOQREPOQREPOQREPOQ 597

RESULT 4

AAB62331
ID AAB62331 standard; Protein: 1162 AA.

XX AAB62331;

DT 29-JUN-2001 (first entry)

DE Amino acid sequence of KSHV tethering protein LANA.

KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;

KM Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;

XX KSHV; latency-associated nuclear antigen; LANA.

OS Kaposi's sarcoma associated herpesvirus.

PN WO200125484-A2.

PD 12-APR-2001.

XX 29-SEP-2000; 2000WO-US26908.

PR 01-OCT-1999; 99US-0410399.

PA (UNMT) UNIV MICHIGAN.

PI Robertson ES, Colter MA;

DR WPI: 2001-281736/29.

PS N-PSDB: AAF62901.

CC A composition for use in gene therapy comprises an expression vector
CC that includes a nucleic acid sequence encoding a nucleic acid binding
CC protein -
CC Disclosure: Fig 9B; 60pp; English.

CC The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for

CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part

PS Claim 1; Page 52; 107pp; English

The sequence represents wheat A-gliadin. A-gliadin derived peptides of the invention are used to test mammalian (preferably human) susceptibility to coeliac disease (gluten intolerance). The peptides are contacted with a blood sample and T cell recognition measured. A positive T-cell recognition indicating a susceptibility to coeliac disease. The peptides are useful for inducing tolerance in an individual and antagonists to the peptides are useful for treating or preventing coeliac disease in an individual and for producing an antibody specific to them or a wild-type sequence. A mutant gliadin protein (or its fragment of 15 amino acids in length) whose wild-type sequence can be modified by transglutaminase to a sequence that comprises the epitope, but which has been modified in such a way that it does not contain sequence which can be modified by transglutaminase to a sequence that comprise the epitope is useful for decreasing the ability of gliadin protein to cause coeliac

CC disease. Nucleic acids encoding proteins antagonistic to the T-cell
CC binding of the epitopes are useful for obtaining a transgenic plant cell
CC or seed and for the production of a protein. The resultant crop plant is
CC useful for obtaining a product of a wheat plant, especially grain, which
CC is optionally processed into flour or another grain product. Food
CC comprising the antagonistic protein is useful instead of a wild-type
CC gliadin.

CC Sequence 266 AA;

Query Match 25.8%; Score 234; DB 22; Length 266;

Best Local Similarity 31.8%; Pred. No. 4,2e-13;

Matches 74; Conservative 15; Mismatches 50; Indels 94; Gaps 11;

QY 3 QLNCSGELGSPQSS---YLDQFYP--QNDYLPQKRFPPYQGFPHPPQYFYLPDELFPQ 57

DB 7 QLDQPNFSQDQPOQVPLVQDQDQFQDQDQFPPQDQFPPQDQFPPQDQFPPQ 63

QY 58 YQIFPPYLPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 111

DB 64 PQLPYP-QPQS-FP--PQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 118

QY 112 ----- 111

DB 119 LQDQILPCMDVYLDQ 178

QY 112 ----PQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 149

DB 179 ILHMQDQ 231

RESULT 9
AAU34519
AAU34519 standard; Protein: 1329 AA.

AC AAU34519;

DT 14-FEB-2002 (first entry)

DE E. coli cellular proliferation protein #100.

KW Antisense: prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS Escherichia coli.

XX WO200170955-A2.

XX 27-SEP-2001.

PD 21-MAR-2001; 2001WO-US09180.

PF 21-MAR-2001; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-25931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;

XX Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

XX N-PSDB: AAS52378.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3: seq ID No 10112; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC tcp.wipo.int/pub/published_pct_sequences.

CC Sequence 1329 AA;

Query Match 24.0%; Score 217; DB 22; Length 1329;

Best Local Similarity 37.3%; Pred. No. 6,7e-11;

Matches 59; Conservative 18; Mismatches 51; Indels 30; Gaps 9;

QY 9 QELQSPQDSTL-QQYPPNPLY-----PQKPF--PVQDQFPHPPQDQFPPY 49

DB 689 QFAOTDQ 746

QY 50 LPDELFPQYQIPPLQDQ 108

DB 747 APQDQ 804

QY 109 PQDQ 141

DB 805 -QDQ 841

RESULT 10

AA844499

ID AA844499 standard; Protein: 308 AA.

AC AA844499;

DT 06-FEB-2001 (first entry)

DE Plant viral movement protein SEQ ID 12.

KW Plant viral movement protein; transport; transgenic plant;

KW viral resistance; cosuppression.

OS Oryza sativa.

XX WO200060088-A2.

XX 12-OCT-2000.

XX 06-APR-2000; 2000WO-US09110.

XX 07-APR-1999; 99US-0128092.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Kriebbers E, Weng Z, Cahoon RE;

XX WPI: 2000-638467/61.

XX N-PSDB: AAC79353.

XX Novel viral movement polypeptides and polynucleotides useful in field

XX of plant molecular biology, for producing transgenic plants, to prepare

XX antibodies and in immunological screening of cDNA expression libraries

XX

FT Region 1331..1373
 FT /label= Exon_XI
 FT Region 1373..1404
 FT /label= Exon_XII
 XX W09213075-A.
 XX
 XX 06-AUG-1992.
 XX
 XX 17-JAN-1992: 92WO-US00433.
 XX
 XX 18-JAN-1991: 91US-0643502.
 PR 10-SEP-1991: 91US-0757022.
 XX
 XX (GENE) GENETICS INST INC.
 XX
 XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K:
 XX
 XX WPI: 1992-284660/34.
 DR N-PSDB: AAQ27223.
 XX
 PT New human mega-karyocyte stimulating factors - for treating
 PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.
 XX
 PS Claim 1, 2 and 3; Fig 1; 87pp; English.
 CC
 CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine preceding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNA's each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor, ie. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.
 XX
 SO Sequence 1404 AA:
 Query Match 22.5%; Score 204; DB 13; Length 1404;
 Best Local Similarity 30.2%; Pred. No. 9.8e-10;
 Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
 QY 12 QSFQOSYLQOQPYPNMPLPQKPFPPVQOQFPHPOOYFPLPEELFPOQIPLPQOQPP 71
 Db 394 KEPAPTTKEPAPTTKEP-APTITKEPAPTTKEP-----PAPTTPKKPAP 445
 QY 72 QOQOQPLP-RPOQFPWQOQOQFPOQOQFPIQOQOQOQOQOQOQOQOQOQOQOQ 130
 Db 446 TTPEKEPAPTTKEPPTTKEPAPTITKEPAPTTKEPAPIAPKKPAPTITKEPAPTTKE 505
 QY 131 SYVQOQOQFPOQO-QPVP 148
 Db 506 PAPTITKEPSPPTTKEPAP 524
 RESULT 13

AAB60568
 ID AAB60568 standard; Protein: 1404 AA.
 XX
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 XX Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
 KM MSF; megakaryocyte stimulating factor; synovial lubricant;
 KM chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KM antiarthritic.
 XX
 OS Homo sapiens.
 OS
 PN W0200107068-A1.
 XX
 PD 01-FEB-2001.
 XX
 XX 21-JUL-2000: 2000WO-US20002.
 PE
 XX 23-JUL-1999: 99US-0145328.
 PR 19-JUL-2000: 2000US-0145328.
 XX
 XX (UYCA-) UNIV CASE WESTERN RESERVE.
 PA
 XX Warman ML;
 PI
 DR WPI: 2001-182721/18.
 XX
 PT New composition comprising the campodactylly-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints
 PT
 XX Example 1: Page -: 34pp; English.
 PS
 XX The invention relates to a method of treating osteoarthritis via the
 XX administration of a composition comprising the campodactylly-arthropathy-
 XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 XX The composition may further comprise a local anesthetic. The composition
 XX of the invention may be administered via intra-articular or intravenous
 XX injection. The human CACP protein is identified in the invention as
 XX being megakaryocyte stimulating factor (MSF). The gene encoding
 XX CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 XX this gene are responsible for the heritable disorder campodactylly-
 XX arthropathy-coxa vara-pericarditis, in which patients have synovial
 XX hyperplasia without evidence of inflammation. CACP protein (MSF)
 XX acts as a synovium lubricant, and can be used to lubricate tissue and
 XX joints in the treatment of osteoarthritis. The composition may be
 XX applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 XX loss of range of movement or joint damage). The present sequence
 XX represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in Figure 4 of the
 CC specification, although a Genbank accession number was given. This
 CC sequence was therefore obtained from Genbank (U070316).
 CC
 SO Sequence 1404 AA:
 Query Match 22.5%; Score 204; DB 22; Length 1404;
 Best Local Similarity 30.2%; Pred. No. 9.8e-10;
 Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
 QY 12 QSFQOSYLQOQPYPNMPLPQKPFPPVQOQFPHPOOYFPLPEELFPOQIPLPQOQPP 71
 Db 394 KEPAPTTKEPAPTTKEP-APTITKEPAPTTKEP-----PAPTTPKKPAP 445
 QY 72 QOQOQPLP-RPOQFPWQOQOQFPOQOQFPIQOQOQOQOQOQOQOQOQOQOQOQ 130
 Db 446 TTPEKEPAPTTKEPPTTKEPAPTITKEPAPTTKEPAPIAPKKPAPTITKEPAPTTKE 505
 QY 131 SYVQOQOQFPOQO-QPVP 148
 Db 506 PAPTITKEPSPPTTKEPAP 524

Wed May 28 09:17:49 2003

us-09-743-533-19.rag

Page 10

[illegible]

Search completed: May 27, 2003, 14:45:50
Job time : 52 secs

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:45:16 : Search time 15 Seconds
(without alignments)

309.921 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906 I MROLNFCSELOSPQOOSYLO.....PFPQOPQVPOQASCTSMKY 158

Sequence: 1 MROLNFCSELOSPQOOSYLO.....PFPQOPQVPOQASCTSMKY 158

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :
1: Issued_patents_AA:*
2: /cgn2_6/p/ptodata/1/iaa/5A.COMB.pep:*
3: /cgn2_6/p/ptodata/1/iaa/5B.COMB.pep:*
4: /cgn2_6/p/ptodata/1/iaa/5A.COMB.pep:*
5: /cgn2_6/p/ptodata/1/iaa/5B.COMB.pep:*
6: /cgn2_6/p/ptodata/1/iaa/5A.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	262.5	29.0	1162	2	US-08-728-323A-2	Sequence 2, Appl1
2	262.5	29.0	1162	2	US-09-298-568-2	Sequence 2, Appl1
3	255.5	28.2	941	4	US-08-991-300-2	Sequence 2, Appl1
4	204	22.5	941	4	US-07-757-022B-14	Sequence 14, Appl1
5	204	22.5	1022	4	US-07-757-022B-84	Sequence 84, Appl1
6	204	22.5	1038	4	US-07-757-022B-74	Sequence 74, Appl1
7	204	22.5	1049	4	US-07-757-022B-58	Sequence 58, Appl1
8	204	22.5	1140	4	US-07-757-022B-104	Sequence 104, Appl1
9	204	22.5	1270	4	US-07-757-022B-44	Sequence 44, Appl1
10	204	22.5	1311	4	US-07-757-022B-42	Sequence 42, Appl1
11	204	22.5	1313	4	US-07-757-022B-142	Sequence 142, Appl1
12	204	22.5	1314	4	US-07-757-022B-50	Sequence 50, Appl1
13	204	22.5	1320	4	US-07-757-022B-46	Sequence 46, Appl1
14	204	22.5	1320	4	US-07-757-022B-60	Sequence 60, Appl1
15	204	22.5	1354	4	US-07-757-022B-48	Sequence 48, Appl1
16	204	22.5	1363	4	US-07-757-022B-40	Sequence 40, Appl1
17	204	22.5	1404	4	US-07-757-022B-52	Sequence 52, Appl1
18	204	22.5	1404	4	US-07-757-022B-2	Sequence 2, Appl1
19	204	22.5	1404	4	US-07-757-022B-62	Sequence 62, Appl1
20	187	20.6	1274	4	US-09-095-443-2	Sequence 2, Appl1
21	184	20.3	667	4	US-08-718-661-2	Sequence 2, Appl1
22	180	19.9	148	4	US-09-178-509-1	Sequence 1, Appl1
23	172	19.0	788	4	US-08-918-914-4	Sequence 4, Appl1
24	169	18.7	605	2	US-08-687-956A-1	Sequence 1, Appl1
25	157.5	17.4	478	3	US-08-153-888-2	Sequence 2, Appl1
26	157.5	17.4	865	4	US-09-281-766-19	Sequence 19, Appl1
27	154	17.0	328	2	US-08-651-818A-2	Sequence 2, Appl1

28	154	17.0	328	4	US-09-184-826-2	Sequence 2, Appl1
29	153	16.9	729	4	US-09-625-188-20	Sequence 20, Appl1
30	150.5	16.6	521	5	PCT-US93-08386-10	Sequence 10, Appl1
31	146.5	16.2	1142	2	US-08-993-118-7	Sequence 7, Appl1
32	146.5	16.2	1142	2	US-08-845-528C-7	Sequence 7, Appl1
33	146.5	16.2	1142	4	US-09-061-709-2	Sequence 2, Appl1
34	146	16.1	2414	4	US-08-227-536-2	Sequence 2, Appl1
35	145.5	16.1	221	4	PCT-US95-04682-2	Sequence 224, App
36	145.5	16.1	221	4	US-09-071-035-224	Sequence 222, App
37	145.5	16.1	282	4	US-09-071-035-222	Sequence 10, Appl1
38	143.5	15.8	312	3	US-08-894-017-10	Sequence 23, Appl1
39	143.5	15.8	1561	3	US-08-894-017-23	Sequence 2, Appl1
40	143.5	15.8	1565	6	US-07-638-431-2	Sequence 2, Appl1
41	142.5	15.7	826	5	PCT-US92-00018-2	Sequence 10, Appl1
42	142.5	15.7	826	5	US-09-281-766-10	Sequence 6, Appl1
43	140.5	15.5	314	2	US-08-525-742-6	Sequence 5, Appl1
44	138	15.2	675	4	US-08-973-273-5	
45	137	15.1				

ALIGNMENTS

RESULT 1
US-08-728-323A-2
Sequence 2, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 05/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 29.0% Score 262.5 DB 2 Length 1162:
Best local similarity 46.8% Pred. No. 5.2e-16:
Matches 74: Conservative 10, Mismatches 51, Indels 23, Gaps 8:
oy 2 ROLNFCSELOSPQOOSYLOQPPVQNPVLPQKPFVQOPQVFPVLPPEEL-FPQVQ 59

[illegible]

```

? RESULT 2
? US-09-298-568-2
? Sequence 2, Application US/09298568
? Patent No. 6322792
? GENERAL INFORMATION:
? APPLICANT: Kieff, Elliott D.
? APPLICANT: Ballester, Mary E.
? APPLICANT: Kaye, Kenneth M.
? TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
? TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
? FILE REFERENCE: 16412-10001R
? CURRENT APPLICATION NUMBER: US/09/298,568
? CURRENT FILING DATE: 1999-04-21
? EARLIER APPLICATION NUMBER: US 60/109,422
? EARLIER FILING DATE: 1998-11-19
? NUMBER OF SEQ ID NOS: 3
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 2
? LENGTH: 1162
? TYPE: prt
? ORGANISM: Kaposi's sarcoma-associated herpesvirus
? US-09-298-568-2

```

	Query Match	29.0%	Score 262.5;	D8 4	Length 1162;
	Best Local Similarity	46.8%;	Pred. No. 5.2e-18;		
	Matches	74;	Conservative	10;	Mismatches 51; Indels 23; Gaps 8
QY	2	ROLNCSGF-IQSFOOSTLOOPYPQNFLPKPFPPVQDPHTTQGYFYLPEEL-FPYOQ	59		
	:	:::::	:	:::	:
D8	453	QQDEFPQOOEPFLQEFPOQDEPQOCFFPQ-----CEPLQDEFQOOEFPQOCPEPLLEPQOQ	502		
QY	60	IPLPLDPPQPFPPO--PQGPLRPQOPFFFWGPPQAPFPOPEPLIHPQDPQPFQPO--PQGP	115		
	:	:::	:	:::	:
D8	503	EFPQOEFPQOOEFPQOOEFPQOOEFPQOOEFPQOOEFPQOOEFPQOOEFPQOREPQOREPQOR	562		
QY	116	FPOQ--PQAIFQOPQOSYPVQ--PQAPFPQOPQPVQ	149		
	:	:::	:	:::	:
D8	563	EPOGREPQ---REFQAREPQAREPQAREPQOREPQOQ	597		

RESULT 3
US-08-991-300-2
Sequence No. 2, Application US/08991300
Patent No. 5973225
GENERAL INFORMATION:
APPLICANT: D'OVIDO, RENATO
APPLICANT: PORCEDDU, ENRICO
APPLICANT: MERCHITELLI, CINZIA
APPLICANT: CARDELLI, LUIGIA EROOLI
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENOME
TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENININ
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
ADDRESSEE: P. C. JEFFERSON DAVIS HIGHWAY
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,300
 FILING DATE: 16-Dec-1997
 CLASSIFICATION: 800
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: IT MI 96/A 002663
 FILING DATE: 19-Dec-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: OBLON, NORMAN F.
 REGISTRATION NUMBER: 24,618
 REESTRAT/DOCKET NUMBER: 2264-0201-0X
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 369 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match	28.2%;	Score 255.5;	DB 2;	Length 369;
Best Local Similarity	41.7%;	Pred. No. 7	1e-18;	
Matches 80;	Conservative 10;	Mismatches 59;	Indels 43;	Gaps 12;

```

QY 3 QLNPPSEGLSPQASVLDLP-----YQNVYLPRK-----PPWQD--PF 4
Db 39 QQQPQSQQQQQQPPLSQQQQQPPPSQQQQQPVLPQQPSPSSQQQLPPFSQQQDPPF 98
QY 41 HPPQQPYFFYLPEE-LPPQYQIPTPLQPPQPPQD-----PQPLRPQPMPQMPQ--CPFP 94
Db 99 SQQQQ--FVLPPQPPSPSQQLPPFSQQLPPFSQQQQPVLPQQPSPSQQQQPPPFSSQLPPFS 156
QY 95 QPQPEI-FQPPQQPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 146
Db 157 QQQQGVLPQQP--PFSQQQQQQLP--PQPPPSQQQQQPVLLQQQLPVRHSLIQQQLNCK 212
QY 149 ---QQASQISM 157
Db 213 VFLQQQCSPPWAM 224

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

FILED DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-14

Query Match 22.5%; Score 204; DB 4; Length 941;
Best Local Similarity 30.2%; Pred. No. 2.3e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSPOOSYLQOPVQONVYLPQKFPVQOPFHTPOQVFPYLPDELFPQOIFPLQPOOPF 71
DB 195 KRPATTTKEPAPTTKPEP-APTTPKEPAPTTKSAPTTPKE-----PAPTTPKRPAP 246
QY 72 QOPQOPLP-RPOQPFVQOPFPOQFPIPOQPOQFPOQFPOQFPOQFPOQFPOQFPOQ 130
DB 247 TTPKEPAPTTKPEPPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 306
QY 131 SYVQPOQOPFPOQF-QPVP 148
DB 307 PAPTTPKEPSPPTTPKPEPAP 325

RESULT 5
US-07-757-022B-84
Sequence 84, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-84

Query Match 22.5%; Score 204; DB 4; Length 1022;
Best Local Similarity 30.2%; Pred. No. 2.5e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSPOOSYLQOPVQONVYLPQKFPVQOPFHTPOQVFPYLPDELFPQOIFPLQPOOPF 71
DB 276 KRPATTTKEPAPTTKPEP-APTTPKEPAPTTKSAPTTPKE-----PAPTTPKRPAP 327
QY 72 QOPQOPLP-RPOQPFVQOPFPOQFPIPOQPOQFPOQFPOQFPOQFPOQFPOQFPOQ 130
DB 328 TTPKEPAPTTKPEPPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 387
QY 131 SYVQPOQOPFPOQF-QPVP 148
DB 388 PAPTTPKEPSPPTTPKPEPAP 406

RESULT 6
US-07-757-022B-74
Sequence 74, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

FILED DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-74

Query Match 22.5% Score 204; DB 4; Length 1038;
Best Local Similarity 30.2%; Pred. No. 2.6e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQGPYPQNPYLQKPEPVQQPFHTPQOYFPYLPBELFPYOYIPTPLQPOQPP 71
DB 260 KEPAATTTKEPAATTTKEP-APTITKEPAATTTKSAPTTPKE-----PAPTTPKKPAP 311
QY 72 QOPQPP-PPQOPFPWQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQ 130
DB 312 TTPKRPATTTKEPTTTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKE 414
QY 131 STPVQPOQPPQOP-QPVP 148
DB 372 PAPTTTKEPSPTTKEPAP 390

RESULT 7
US-07-757-022B-58
Sequence 58, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-58

Query Match 22.5% Score 204; DB 4; Length 1049;
Best Local Similarity 30.2%; Pred. No. 2.6e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQGPYPQNPYLQKPEPVQQPFHTPQOYFPYLPBELFPYOYIPTPLQPOQPP 71
DB 303 KEPAATTTKEPAATTTKEP-APTITKEPAATTTKSAPTTPKE-----PAPTTPKKPAP 354
QY 72 QOPQPP-PPQOPFPWQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQ 130
DB 355 TTPKRPATTTKEPTTTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKE 414
QY 131 STPVQPOQPPQOP-QPVP 148
DB 415 PAPTTTKEPSPTTKEPAP 433

RESULT 8
US-07-757-022B-104
Sequence 104, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: AMINO ACID
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-07-757-022B-104

Query Match 22.5%; Score 204; DB 4; Length 1140;
Best Local Similarity 30.2%; Pred. No. 2.9e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4:

QY 12 QSPQOSYLQOPYPQNPYLPQKPFVQOPFHPTPOQYFPLPEELFPQYQIPTPLQPOOPFP 71
DB 394 KEPAPTTKKEPAFTTKKEP-AFTTKKEPAFTTKSAFTTKKE-----PAPTTKKRPAP 445
QY 72 QDPQQLP-RPQPFPMQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 130
DB 446 TPKEPAFTTKKEPFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKKE 505
QY 131 SYFVQOPFPQOP-QPVP 148
DB 506 PAPTTTKESPPTTKKEPAP 524

RESULT 9
US-07-757-022B-44
Sequence 44, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: AMINO ACID
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-07-757-022B-44

Query Match 22.5%; Score 204; DB 4; Length 1270;
Best Local Similarity 30.2%; Pred. No. 3.2e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4:

QY 12 QSPQOSYLQOPYPQNPYLPQKPFVQOPFHPTPOQYFPLPEELFPQYQIPTPLQPOOPFP 71
DB 260 KEPAPTTKKEPAFTTKKEP-AFTTKKEPAFTTKSAFTTKKE-----PAPTTKKRPAP 311
QY 72 QDPQQLP-RPQPFPMQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 130
DB 312 TPKEPAFTTKKEPFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKKE 371
QY 131 SYFVQOPFPQOP-QPVP 148
DB 372 PAPTTTKESPPTTKKEPAP 390

RESULT 10
US-07-757-022B-42
Sequence 42, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196

RESULT 12
 US-07-757-022B-50
 Sequence 50, Application US/07757022B
 Patent No. 6433142
 GENERAL INFORMATION:
 APPLICANT: Genent, Thomas G.
 APPLICANT: Clark, Stephen C.
 APPLICANT: Turner, Katherine
 APPLICANT: Hewick, Rodney M.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Csefi, Luann

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 22.5%; Score 204; DB 4; Length 1314;
Best Local Similarity 30.2%; Pred. No. 3,3e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
QY 12 QSPQSYLQDPYDNPYLPQKPPVQOPPHPPQOYFPLYLPELFPQYQIPIPLQOPQOPFP 71
DB 304 KEPAPTTKKEPAFTTKREP-APTTKKEPAFTTKSAPTTPKE-----PAPTTPKKKBP 355
QY 72 QQPQOPPLP-RPQOPFPWQOPQOPQEP1PQOPQOPFPQOPQOPFPQOPQOII1PQOPQO 130
DB 356 TTPKEPAFTTKKEPTTPKEPAFTTKKEPAFTTKKEPAFTAKKPAFTTKKEPAFTTPKE 415
QY 131 SYPVQPOQOPFPQOP-QVP 148
DB 416 PAPTTPKEPTTKKEPA 434

RESULT 13
US-07-757-022B-46
Sequence 46, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-46

Query Match 22.5%; Score 204; DB 4; Length 1320;
Best Local Similarity 30.2%; Pred. No. 3,4e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
QY 12 QSPQSYLQDPYDNPYLPQKPPVQOPPHPPQOYFPLYLPELFPQYQIPIPLQOPQOPFP 71
DB 310 KEPAFTTKKEPAFTTKREP-APTTKKEPAFTTKSAPTTPKE-----PAPTTPKKKBP 361
QY 72 QQPQOPPLP-RPQOPFPWQOPQOPQEP1PQOPQOPFPQOPQOPFPQOPQOII1PQOPQO 130
DB 362 TTPKEPAFTTKKEPTTPKEPAFTTKKEPAFTTKKEPAFTAKKPAFTTKKEPAFTTPKE 421
QY 131 SYPVQPOQOPFPQOP-QVP 148
DB 422 PAPTTPKEPTTKKEPA 440

RESULT 14
US-07-757-022B-60
Sequence 60, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
us-07-757-022B-60

Query Match 22.5%; Score 204; DB 4; Length 1320;
Best Local Similarity 30.2%; Pred. No. 3.4e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

Db 12 QSPQSYLQOPYPQNPVLPQKPFVQOPFHTPQGFYLYLEELFPOYQITPLQPOQPP 71
Db 310 KRPAPTTTKRPAPTTKRP-APTTRKRPAPTTTKSAPTTRK-----PAPTTKRPAP 361
QY 72 QOPQOLP-RPQOPFQWQOPFQOPFQOPFQOPFQOPFQOPFQOPFQOPFQOPFQOP 130
Db 362 TTRKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTK 421
QY 131 SYVQOPQOPFQOP-OPVP 148
Db 422 PAPTTTKRPSPPTTKRPAP 440

RESULT 15
US-07-757-022B-48
Sequence 48, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caser, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
us-07-757-022B-48

Query Match 22.5%; Score 204; DB 4; Length 1354;
Best Local Similarity 30.2%; Pred. No. 3.5e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

Db 12 QSPQSYLQOPYPQNPVLPQKPFVQOPFHTPQGFYLYLEELFPOYQITPLQPOQPP 71
Db 344 KRPAPTTTKRPAPTTKRP-APTTRKRPAPTTTKSAPTTRK-----PAPTTKRPAP 395
QY 72 QOPQOLP-RPQOPFQWQOPFQOPFQOPFQOPFQOPFQOPFQOPFQOPFQOPFQOP 130
Db 396 TTRKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTKRPAPTTK 455
QY 131 SYVQOPQOPFQOP-OPVP 148
Db 456 PAPTTTKRPSPPTTKRPAP 474

Search completed: May 27, 2003, 14:48:11
Job time: 26 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:45:55 : Search time 21 Seconds

(without alignments)
746.156 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906 I MRQLNPGSOFLOSPQDSYD.....PPPOQPPVPOQASCTSMV 158

Sequence: 1 MRQLNPGSOFLOSPQDSYD.....PPPOQPPVPOQASCTSMV 158

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications-AA:*

1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppa/PTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244.5	27.0	1362	10	US-09-815-242-14009
2	222	24.5	296	9	US-09-789-054A-10
3	217	24.0	1329	10	US-09-815-242-10112
4	204	22.5	941	12	US-10-124-557-14
5	204	22.5	1022	12	US-10-124-557-84
6	204	22.5	1038	12	US-10-124-557-74
7	204	22.5	1049	12	US-10-124-557-58
8	204	22.5	1140	12	US-10-124-557-104
9	204	22.5	1270	12	US-10-124-557-44
10	204	22.5	1311	12	US-10-124-557-42
11	204	22.5	1313	12	US-10-124-557-142
12	204	22.5	1314	12	US-10-124-557-50
13	204	22.5	1320	12	US-10-124-557-46
14	204	22.5	1320	12	US-10-124-557-60
15	204	22.5	1334	12	US-10-124-557-48
16	204	22.5	1361	12	US-10-124-557-40
17	204	22.5	1363	12	US-10-124-557-52
18	204	22.5	1404	12	US-10-124-557-2
19	204	22.5	1404	12	US-10-124-557-62

20	189	20.9	312	9	US-09-789-054A-22	Sequence 22, App1
21	187	20.6	1274	9	US-10-020-215-2	Sequence 2, App1
22	183.5	20.3	371	9	US-09-284-320-2	Sequence 2, App1
23	183.5	20.3	371	9	US-10-197-666A-4	Sequence 4, App1
24	180	19.9	148	12	US-10-117-121-1	Sequence 1, App1
25	171.5	18.9	254	10	US-09-216-393-126	Sequence 126, App
26	167.5	18.5	171	10	US-09-216-393-71	Sequence 71, App1
27	167.5	18.5	171	10	US-09-216-393-274	Sequence 274, App
28	164.5	18.2	171	9	US-10-077-584-4	Sequence 4, App1
29	163.5	18.0	503	12	US-10-062-254-240	Sequence 240, App
30	161.5	17.8	340	9	US-09-789-054A-6	Sequence 6, App1
31	157.5	17.4	199	9	US-10-034-934-125	Sequence 125, App
32	157.5	17.4	865	9	US-09-957-995A-19	Sequence 19, App1
33	157	17.3	348	9	US-09-788-054A-20	Sequence 20, App1
34	156.5	17.3	2005	10	US-09-735-367B-3	Sequence 3, App1
35	156.5	17.3	2063	10	US-09-735-367B-2	Sequence 2, App1
36	156	17.2	370	9	US-10-021-955-78	Sequence 78, App1
37	156	17.2	370	9	US-10-021-955-83	Sequence 83, App1
38	156	17.2	370	9	US-10-021-955-90	Sequence 90, App1
39	155.5	17.2	370	9	US-10-021-955-91	Sequence 91, App1
40	154	17.0	407	9	US-09-738-626-5433	Sequence 5433, Ap
41	153	16.9	666	10	US-09-791-171-70	Sequence 70, App1
42	152	16.8	668	10	US-09-925-300-1633	Sequence 1633, Ap
43	151	16.7	868	10	US-09-800-729-106	Sequence 106, App
44	151	16.7	921	10	US-09-800-729-199	Sequence 199, App
45	150.5	16.6	356	10	US-09-735-367B-13	Sequence 13, App1

ALIGNMENTS

RESULT 1
US-09-815-242-14009
Sequence 14009 Application US/09815242
Patient No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl F.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/151,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 14009
LENGTH: 1362
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(1362)
OTHER INFORMATION: Xaa = Any Amino Acid

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Query Match 22.5%; Score 204; DB 12; Length 941;
Best Local Similarity 30.2%; Pred. No. 6e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQPPYLPQKPFPPVQOFPHPYLPPELFPQYQIPTPLQPOQPP 71
DB 195 KEPAPTTKPEAPPTPKPE-APTTKPEAPPTTKSAPTTKPE-----PAPTTPKKPAP 246
QY 72 QGQOQPLP-RPOQFPPOQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 130
DB 247 TTKEPAPTTPKEPTTKPEAPTTKEPAPTTPKEAPTTAKKAPPTTKPEAPTTKE 306
QY 131 STYVQPOQPPPOQ-OPVP 148
DB 307 PAPTTTKPEPPTTKPEPAP 325

RESULT 5

US-10-124-557-84
Sequence 84, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match 22.5%; Score 204; DB 12; Length 1022;
Best Local Similarity 30.2%; Pred. No. 6.5e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQPPYLPQKPFPPVQOFPHPYLPPELFPQYQIPTPLQPOQPP 71
DB 276 KEPAPTTKPEAPPTPKPE-APTTKPEAPPTTKSAPTTKPE-----PAPTTPKKPAP 327
QY 72 QGQOQPLP-RPOQFPPOQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 130
DB 328 TTKEPAPTTPKEPTTKPEAPTTKEPAPTTPKEAPTTAKKAPPTTKPEAPTTKE 387
QY 131 STYVQPOQPPPOQ-OPVP 148
DB 388 PAPTTTKPEPPTTKPEPAP 406

RESULT 6

US-10-124-557-74
Sequence 74, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 22.5% Score 204; DB 12; Length 1038;
Best Local Similarity 30.2%; Pred No. 6.6e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQOSTIQQPYQNPYLPQKPFVQQPPHTPQOYFPLYPELFPYOIPTPLQPOQPP 71
DB 260 KEPAPTTKKEPAPTTKPEP-APTTTKKEPAPTTKSAPTTKPE-----PAPTPKRPAP 311
OY 72 QOPQOPLP-RQOPFPWQPOQPFPOEPPIPOQPOQPFPOQPOQPFPOQOITFOQPOQ 130
DB 312 TTPKEPAPTTKKEPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 371
OY 131 SYVQPOQPPPOQ-OPVP 148
DB 372 PAPTTTKESPTTKPEPAP 390

RESULT 7
US-10-124-557-58
Sequence 58, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 22.5% Score 204; DB 12; Length 1049;
Best Local Similarity 30.2%; Pred No. 6.7e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQOSTIQQPYQNPYLPQKPFVQQPPHTPQOYFPLYPELFPYOIPTPLQPOQPP 71
DB 303 KEPAPTTKKEPAPTTKPEP-APTTTKKEPAPTTKSAPTTKPE-----PAPTPKRPAP 354
OY 72 QOPQOPLP-RQOPFPWQPOQPFPOEPPIPOQPOQPFPOQPOQPFPOQOITFOQPOQ 130
DB 355 TTPKEPAPTTKKEPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 414
OY 131 SYVQPOQPPPOQ-OPVP 148
DB 415 PAPTTTKESPTTKPEPAP 433

RESULT 8
US-10-124-557-104
Sequence 104, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match
Best Local Similarity 30.2%; Pred. No. 7,2e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

22.5%; Score 204; DB 12; Length 1140;
DB 12 QSPQSYLQQYPQNPVLPQKFPVQOPFHTPQOYFPLYPEELFPQYOIPTPLQPOQFP 71
DB 394 KEPAPTTKEPAPPTPKPE-APTPTKEPAPTTKSAPTTKE-----PAPTTKKRAP 445
QY 72 QCPQQLP-RPQOPFPWQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFP 130
DB 446 TTKKEPAPTTKEPPTTKKEPAPTTKEPAPTTKEPAPTTKAPKRAVTTTKEPAPTTKE 505
QY 131 SYVQPOQOPFPQOP-QPVP 148
DB 506 PAPTTKEPSPPTTKEPAP 524

RESULT 9
US-10-124-557-44
Sequence 44, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Gesner, Rodney M.
Hewick, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserf, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

Query Match
Best Local Similarity 30.2%; Pred. No. 8e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

22.5%; Score 204; DB 12; Length 1270;
DB 12 QSPQSYLQQYPQNPVLPQKFPVQOPFHTPQOYFPLYPEELFPQYOIPTPLQPOQFP 71
DB 260 KEPAPTTKEPAPPTPKPE-APTPTKEPAPTTKSAPTTKE-----PAPTTKKRAP 311
QY 72 QCPQQLP-RPQOPFPWQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFP 130
DB 312 TTKKEPAPTTKEPPTTKKEPAPTTKEPAPTTKEPAPTTKAPKRAVTTTKEPAPTTKE 371
QY 131 SYVQPOQOPFPQOP-QPVP 148
DB 372 PAPTTKEPSPPTTKEPAP 390

RESULT 10
US-10-124-557-42
Sequence 42, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Gesner, Rodney M.
Hewick, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserf, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match 22.5%: Score 204; DB 12; Length 1311;
Best Local Similarity 30.2%; Pred. No. 8.3e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 OSPOOSYLOOPYPQNPVLPQKFPVQDPFHTPQOYFPLDELFPQYQIPTPLQPOQPP 71
DB 301 KEPAFTTKKEPAPTTKPEP-APTTTKEPAPTTTKSAPTTTKE-----PAPTTTKEPAP 352
OY 72 QOPQOPLP-RPQOPFPWQOPFPQOPQEPITPQOPQOPFPQOPFPQOPQOIIIFQOPQO 130
DB 353 TTPKEPAPTTTKEPPTTTPKEPAPTTTKEPAPTTTKEPAPTAAPKRPAPTTTKEPAPTTTKE 412
OY 131 SYVQOPQOPFPQOP-OPVP 148
DB 413 PAPTTTKEPSPTTTKEPAP 431

RESULT 11

US-10-124-557-142
Sequence 142, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Caserf, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match 22.5%: Score 204; DB 12; Length 1313;
Best Local Similarity 30.2%; Pred. No. 8.3e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 OSPOOSYLOOPYPQNPVLPQKFPVQDPFHTPQOYFPLDELFPQYQIPTPLQPOQPP 71
DB 303 KEPAFTTKKEPAPTTKPEP-APTTTKEPAPTTTKSAPTTTKE-----PAPTTTKEPAP 354
OY 72 QOPQOPLP-RPQOPFPWQOPFPQOPQEPITPQOPQOPFPQOPFPQOPQOIIIFQOPQO 130
DB 355 TTPKEPAPTTTKEPPTTTPKEPAPTTTKEPAPTTTKEPAPTAAPKRPAPTTTKEPAPTTTKE 414
OY 131 SYVQOPQOPFPQOP-OPVP 148
DB 415 PAPTTTKEPSPTTTKEPAP 433

RESULT 12

US-10-124-557-50
Sequence 50, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Caserf, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:

US-10-124-557-50

Query Match

Best Local Similarity 22.5%; Score 204; DB 12; Length 1314;
30.2%; Pred. No. 8.3e-08;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSPOSVLQOPYPONRYLPKPFPPVOQPFPHPPQOYFPLPELFPQYQIPLPQOQPPF 71

DB 304 KEPAFTTKREPAPTTKPEP-APTTRKEPAPTTTKSAPTTTKE-----PAPTTTKKPPAP 355

QY 72 QOPQOPLP-RPOQPFPPWOQOQPFPOQFPIPOQPOQPFPPQOQPOQIIFQOQPOQ 130

DB 356 TTRKEPAPTTKPEPPTTRKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKE 415

QY 131 SYVPOQOPFPPOOP-OPVP 148

DB 416 PAPTTTKESPPTTKPEPAP 434

RESULT 13

US-10-124-557-46

Sequence 46, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserf, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 46:

US-10-124-557-46

Query Match

22.5%; Score 204; DB 12; Length 1320;

Best Local Similarity 30.2%; Pred. No. 8.3e-08;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

Best Local Similarity 30.2%; Pred. No. 8.3e-08;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSPOSVLQOPYPONRYLPKPFPPVOQPFPHPPQOYFPLPELFPQYQIPLPQOQPPF 71

DB 310 KEPAFTTKREPAPTTKPEP-APTTRKEPAPTTTKSAPTTTKE-----PAPTTTKKPPAP 361

QY 72 QOPQOPLP-RPOQPFPPWOQOQPFPOQFPIPOQPOQPFPPQOQPOQIIFQOQPOQ 130

DB 362 TTRKEPAPTTKPEPPTTRKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKE 421

QY 131 SYVPOQOPFPPOOP-OPVP 148

DB 422 PAPTTTKESPPTTKPEPAP 440

RESULT 14

US-10-124-557-60

Sequence 60, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserf, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 60:

US-10-124-557-60

Query Match

22.5%; Score 204; DB 12; Length 1320;

Best Local Similarity 30.2%; Pred. No. 8.3e-08;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQSYLQPPYRQNFYLPQKPFVQOPFHTPQOYFPLLEBELFPOXQIPPLQPOQPP 71
 Db 310 KEPAFTTKKEPAFTTKPE-APTTTKEPAFTTKSAFTTKE-----PAPTTKPPAP 361
 OY 72 QOPQPLP-RPQOPPPMQPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 130
 Db 362 TTKKEPAFTTKPEFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKE 421
 OY 131 SYVQOPQOPFPQOP-QPVP 148
 Db 422 PAPTTTKEPSPTTKPEAP 440

RESULT 15

US-10-124-557-48

: Sequence 48, Application US/10124557

: Patent No. US20020137894A1

: GENERAL INFORMATION:

: APPLICANT: Turner, Katherine
 : Clark, Stephen C.
 : Jacobs, Kenneth
 : Hewick, Rodney M.
 : Gesner, Thomas G.
 : TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 : NUMBER OF SEQUENCES: 143
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Genetics Institute, Inc.
 : STREET: 87 Cambridgepark Drive
 : CITY: Cambridge
 : STATE: Massachusetts
 : COUNTRY: U.S.A.
 : ZIP: 02140

: COMPUTER READABLE FORM:

: MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/10/124,557
 : FILING DATE: 16-Apr-2002
 : CLASSIFICATION: <unknown>
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/643,502
 : FILING DATE: 18-Jan-1991
 : APPLICATION NUMBER: US 07/546,114
 : FILING DATE: 29-Jun-1990
 : APPLICATION NUMBER: US 07/457,196
 : FILING DATE: 29-Dec-1989
 : APPLICATION NUMBER: US 07/390,901
 : FILING DATE: 08-Aug-1989
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Casert, Luann
 : REGISTRATION NUMBER: 31,822
 : REFERENCE/DOCKET NUMBER: GI 5190
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617)876-1170
 : TELEFAX: (617)876-5851
 : INFORMATION FOR SEQ ID NO: 48:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1354 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : SEQUENCE DESCRIPTION: SEQ ID NO: 48:
 : US-10-124-557-48

Query Match 22.5%; Score 204; DB 12; Length 1354;
 Best Local Similarity 30.2%; Pred. No. 8.5e-08;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQSYLQPPYRQNFYLPQKPFVQOPFHTPQOYFPLLEBELFPOXQIPPLQPOQPP 71
 Db 344 KEPAFTTKKEPAFTTKPE-APTTTKEPAFTTKSAFTTKE-----PAPTTKPPAP 395

OY 72 QOPQPLP-RPQOPPPMQPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 130
 Db 396 TTKKEPAFTTKPEFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKKEPAFTTKE 455
 OY 131 SYVQOPQOPFPQOP-QPVP 148
 Db 456 PAPTTTKEPSPTTKPEAP 474

Search completed: May 27, 2003, 14:48:34
 Job time : 27 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:36:30 : Search time 15 Seconds
(without alignments)

436.884 Million cell updates/sec

Title: US-09-743-533-19

Sequence: 1 MRLNPGSGLQSPQSYLQ.....PFPQOPQVPOQASCTSMW 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	412	45.5	327	1	GDB2_WHEAT
2	384.5	42.4	302	1	GDBX_WHEAT
3	347	38.3	251	1	GDBO_WHEAT
4	332	36.6	291	1	GDBB_WHEAT
5	310.5	34.3	105	1	HOR1_HORVU
6	302	33.3	305	1	HOG1_HORVU
7	289.5	32.0	307	1	GDA5_WHEAT
8	272.5	30.1	72	1	HOR8_HORVU
9	271.5	30.0	293	1	HOR1_HORVU
10	267.5	29.5	289	1	HOG3_HORVU
11	263.5	29.1	297	1	GDA4_WHEAT
12	255	28.1	296	1	GDA6_WHEAT
13	250	27.6	291	1	GDA2_WHEAT
14	242	26.7	282	1	GDA3_WHEAT
15	240	26.5	264	1	HOR3_HORVU
16	238	26.3	356	1	GITA_WHEAT
17	236.5	26.1	319	1	GDA7_WHEAT
18	236	26.0	313	1	GDA7_WHEAT
19	234	25.8	286	1	GDA0_WHEAT
20	232.5	25.7	262	1	GDA1_WHEAT
21	218.5	24.1	68	1	HOR9_HORVU
22	217	24.0	1329	1	FTSB_ECOLI
23	211.5	23.3	307	1	GLTB_WHEAT
24	208.5	23.0	462	1	ANX7_DICDI
25	206	22.7	213	1	AMEX_BOVIN
26	205	22.6	202	1	AMEL_MONDO
27	202	22.3	1902	1	SMF1_HUMAN
28	201.5	22.2	152	1	CORA_RAT
29	196	21.6	467	1	CBPA_DICDI
30	195	21.5	2364	1	PCGA_BOVIN
31	189	20.9	126	1	CORA_RABIT
32	189	20.8	1362	1	BRD4_HUMAN
33	188.5	20.8	144	1	CORA_MOUSE

34	186	20.5	191	1	AMEX_HUMAN
35	182.5	20.1	555	1	Q9F96_chi1amydomon
36	182	20.1	738	1	YKFA_YEAST
37	181.5	20.0	634	1	HMP1_CANAL
38	181	20.0	255	1	LP61_ETMTE
39	180	19.9	1048	1	SR4A_RAT
40	177.5	19.6	196	1	AMEL_MOUSE
41	177	19.5	304	1	GDB1_WHEAT
42	176	19.4	746	1	PCAP_HUMAN
43	175.5	19.4	875	1	Y066_NPYOP
44	175	19.3	3164	1	TEGU_HSV1
45	170.5	18.8	244	1	GDB3_WHEAT
					P04730_triticum ae

ALIGNMENTS

```

RESULT 1
GDB2_WHEAT
ID GDB2_WHEAT STANDARD: PRT: 327 AA.
AC P08453:
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Gamma-gliadin precursor.
OS Triticum aestivum (wheat).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae:
OC Triticeae: Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiyama T., Rafalski A., Soell D.;
RT "The nucleotide sequence of a wheat gamma-gliadin genomic clone.";
RL Plant Sci. 44:205-209(1986).
CC -i- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -i- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC NEAR DIRECT REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isd-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M16064; AAA34289.1; .
CC PIR: JS0402; JS0402.
CC InterPro: IPR003612; AAI.
CC InterPro: IPR001954; G1a-glutenin.
CC InterPro: IPR001768; try/amy1_inhbr.
CC Pfam: PF00234; try_alpha_amy1.1.
CC PRINTS: PR00206; GLIADGLUTEN.
CC SMART: SM00499; AAI; 1.
CC Seed storage protein; Repeat; Signal; Multigene family.
CC STGNL 1 327 GAMMA-GLIADIN.
CC FT CHAIN 20 327
CC SO SEQUENCE 327 AA: 37122 MW: E27FE9DABDFCCB CRC64;

Query Match 45.5%; Score 412; DB 1; Length 327;
Best local Similarity 50.0%; Pred. No. 3.9e-22;
Matches 97; Conservative 12; Mismatches 37; Indels 48; Gaps 10;

QY 3 QLNCSGE--LQSPQSYLQOPYPONFYLPQKFPYQOPFPFPQYFPLYPEELPFOYQI 60
Db 22 QVDSGGVQWLOQQLVQLOQPLSQP-----QQTFFQYQYFPFPQYQVQVFPQ 72
QY 61 PTP--LQPPQFPFPQPPQPLP---RPQQFPFPQPPQPP---OPQFFIPQPPQPPFP--Q 110
Db 73 PQQFFLPQPPQFPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 132

```

```

QY 111 OPOOPPP--OOPQOIFPOOPQOS--YVPOPOPPPOQPPVPOQA----- 151
Db 133 QPOOPFPQOQQOPOQ--FPPOPOQOLPOPOQPOQOSFPQOQPFQIOPSLQOOLNPKNILLQO 191
QY 152 -----SCISMW 158
Db 192 SKPASTVSLMSIT 205

RESULT 2
GDBX_WHEAT
ID GDBX_WHEAT STANDARD: PRT: 302 AA.
AC P21292;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE Gamma-gliadin precursor.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN 11]
RP SEQUENCE FROM N.A.
RC SRRAIN-cv. Yamhill;
RA Scheets K., Hedgcock C.;
RT "Nucleotide sequence of a gamma gliadin gene: comparisons with other
RT gamma gliadin sequences show the structure of gamma gliadin genes and
RT the general primary structure of gamma gliadins."
RL Plant Sci. 57:141-150(1988).
CC 1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC 1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC NEAR DIRECT REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M36999; AAA34272.1; .
CC PIR: JAO153; JAO153.
CC InterPro: IPR003612; AAI.
CC InterPro: IPR001954; Glia_glu1enin.
CC InterPro: IPR001768; Try/amyL_inh1tr.
CC Pfam: PF00234; try_alpha_amyL; 1.
CC PRINTS: PR00208; GLIADGLUTEN.
CC SMART: SM00499; AAI; 1.
CC Seed storage protein; Repeat; Signal; Multigene family.
CC SIGNAL 1 19
CC CHAIN 20 302 GAMMA-GLIADIN.
CC SEQUENCE 302 AA; 34300 MW; 672437C770D6F82 CRC64;

Query Match 42.4%; Score 384.5; DB 1; Length 302;
Best Local Similarity 58.0%; Pred. No. 2.8e-20;
Matches 91; Conservative 11; Mismatches 28; Indels 27; Gaps 11;

```

```

RESULT 3
GDB0_WHEAT
ID GDB0_WHEAT STANDARD: PRT: 251 AA.
AC P08079;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Gamma-gliadin precursor (Fragment).
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN 11]
RP SEQUENCE FROM N.A.
RA Scheets K., Rafalski J.A., Hedgcock C., Soell D.G.;
RT "Heptapeptide repeat structure of a wheat gamma-gliadin."
RL Plant Sci. Lett. 37:221-225(1985).
CC 1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC 1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC NEAR DIRECT REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M10600; AAA34288.1; .
CC PIR: PS0094; PS0094.
CC InterPro: IPR001768; Try/amyL_inh1tr.
CC Pfam: PF00234; try_alpha_amyL; 1.
CC Seed storage protein; Repeat; Signal; Multigene family.
CC SIGNAL 1 19
CC CHAIN 20 >251 GAMMA-GLIADIN.
CC FT NON-TER 251 251
CC SEQUENCE 251 AA; 29054 MW; 5A3F12C36C825EAD CRC64;

Query Match 38.3%; Score 347; DB 1; Length 251;
Best Local Similarity 45.0%; Pred. No. 8.7e-18;
Matches 86; Conservative 11; Mismatches 24; Indels 70; Gaps 11;

```

OS *Triticum aestivum* (Wheat).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 CC Triticeae; Triticum.
 CC NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86301876; PubMed=3017812;
 RA Rafalski J.A.;
 RT "Structure of wheat gamma-gliadin genes";
 RL Gene 43:221-229(1986).
 CC -I- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -I- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
 CC NEAR DIRECT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M13713; AAA34274.1;
 CC PIR: A25632; EEWG.
 CC InterPro: IPR003612; ANI.
 CC InterPro: IPR001954; G1a_gluTenin.
 CC InterPro: IPR001768; try/amy1_inhbr.
 CC Pfam: PF00234; tryP_alpha_amy1.1.
 CC PRINTS: PR00208; GLIADGLUTEN.
 CC SMART: SM00499; AAI: 1.
 CC Seed storage protein; Repeat; Signal; Multigene family.
 FT SIGNAL 1 19
 FT CHAIN 20 291 GAMMA-GLIADIN B.
 FT SEQUENCE 291 AA: 32967 MW: 85732222A2EFA01 CRC64;
 SQ
 Query Match 36.6%; Score 332; DB 1; Length 291;
 Best Local Similarity 55.2%; Pred. No. 1.1e-16;
 Matches 69; Conservative 9; Mismatches 23; Indels 24; Gaps 5;
 QY 31 QKPPPPQPFHPPQPFYLLPEELFPQYQITPTPLQPPQPPFPQPPQPPFPQPPQ 90
 DB 28 QVQWPPQOQFFLQHPQFQSOQPPQIFP-----QPPQTPHPQOQDFPQ-----QPP 74
 QY 91 QPPPPQPFHPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 146
 DB 75 QQFLQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 128
 QY 147 -VPPQ 150
 DB 129 SPFQ 133
 RESULT 5
 HOG1_HORVU
 ID HOG1_HORVU STANDARD: PRT: 105 AA.
 AC P06472;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE C-hordein (PCP387) (Fragment).
 OS Hordeum vulgare (Barley).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 CC Triticeae; Hordeum.
 CC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Forde B.G., Kreis M., Williamson M.S., Fry R.P., Pywell J.,
 RA Shewry P.R., Bunce N., Milfin B.J.;
 RT "Short tandem repeats shared by B- and C-hordein cDNAs suggest a

RT common evolutionary origin for two groups of cereal storage protein
 RT genes.";
 RL EMBO J. 4:9-15(1985).
 CC -I- FUNCTION: SULFUR-POOR SEED STORAGE PROTEIN.
 CC -I- TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X01779; CAA25914.1;
 CC PIR: S07189; S07189.
 CC Seed storage protein; Multigene family.
 FT NON_TER 1
 FT SEQUENCE 105 AA: 12180 MW: 3845823443652F CRC64;
 SQ
 Query Match 34.3%; Score 310.5; DB 1; Length 105;
 Best Local Similarity 50.3%; Pred. No. 1.3e-15;
 Matches 73; Conservative 8; Mismatches 23; Indels 41; Gaps 7;
 QY 14 PQQSVLQPPYPPQPPYLPQKPPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 73
 DB 2 PQQSV-----PVPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 40
 QY 74 PQQPLPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 133
 DB 41 SEQLI--PQQPPFLQPPQ-----PFPQPPQPPQPPQPPQPPQPPQPPQPPQPP 89
 QY 134 VQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 158
 DB 90 LQPHQPPYQQ-----TIMSMV 105
 RESULT 6
 HOG1_HORVU
 ID HOG1_HORVU STANDARD: PRT: 305 AA.
 AC P17990;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Gamma-hordein 1 precursor.
 OS Hordeum vulgare (Barley).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 CC Triticeae; Hordeum.
 CC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=cv. Carina;
 RA Cameron-Mills V., Brandt A.;
 RT "A gamma-hordein gene";
 RL Plant Mol. Biol. 11:449-461(1988).
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC (AS GLOBULES) AND VACUOLAR
 CC (AS PROTEIN BODIES).
 CC -I- TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.
 CC -I- DOMAIN: SULFUR RICH HORDEIN WHICH POSSES AN N-TERMINAL HALF
 CC COMPOSED OF PROLINE-GLUTAMINE BLOCKS ORGANIZED IN REPEATING UNITS
 CC AND A C-TERMINAL HALF WHERE THE REPEATS ARE DISPERSED AND LESS
 CC CONSERVED.
 CC -I- SIMILARITY: TO GAMMA-GLIADIN FROM WHEAT, AND A GAMMA-SECALIN FROM
 CC RYE, AND LESS TO A B1 HORDEIN FROM BARLEY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DE Alpha/beta-gliadin A-II precursor (Prolamin).
 OS Triticum aestivum (Wheat).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 CC Triticaceae; Triticum.
 CC NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=65234522; PubMed=2989281;
 RA Okita T.W., Chesbrough V., Reeves C.D.;
 RT Evolution and heterogeneity of the alpha-/beta-type and gamma-type
 RT gliadin DNA sequences.";
 RJ J. Biol. Chem. 260:8203-8213(1985).
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -1- MISCELLANEOUS: THE ALPHA/BETA-GLIADIN CAN BE DIVIDED INTO 5
 CC HOMOLOGY BASES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE
 CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
 CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
 CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: M1076; AAA34283.1; -
 DR PIR: E22364; E22364.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001954; Glla_gluTenH.
 DR InterPro: IPR001768; TRY/amy1_inhtr.
 DR Pfam: PF00234; tryf-alpha_amy1.1.
 DR PRINTS: PR00208; GLIADGLUTEN.
 DR SMART: SM00499; AAI; 1.
 DR SMap storage protein: Repeat; Signal; Multigene family.
 FT SIGNAL 1
 FT CHAIN 21 282 ALPHABETA-GLIADIN A-II;
 FT SEQUENCE 282 AA; 32236 MW; 87277047784E870A CRC64;
 Query Match 26.7%; Score 242; DB 1; Length 282;
 Best Local Similarity 39.7%; Pred. NO. 1.5e-10;
 Matches 69; Conservative 16; Mismatches 49; Indels 40; Gaps 10.
 Oy 3 QLNPCSOELSPPOOS---YLQGPVP-QNPYLQKFFPVQGGFFHTPOQFPYLPDELFP 56
 Db 27 QLOPQNPSSQOPOEPOVPLMQOQOQFFGQOEQFPQGPYPHQGPFFSOPPY-----P 77
 Oy 57 QVOETPLPQOPOPPPOQPOQPLRPPOQPPMPOQPPPOQCEPIPOQPOQPPPOQPOPF 116
 Db 78 Q--PQPPPPQPPVPO--TGPFP-PQGPVP-QPQPPYPPQGPQISQQAQQAQ--QDQDQPTL 129
 Oy 117 PQQPPQ-----TLPQPP-----QOSYVQPPQPPQPPPPVPOQAQSC 153
 Db 130 QQLIQQLIPCRDQVLDQHNHIAASSQVLDQSSVQQLDQLCCQLLPQIPQSRIC 183
 RESULT 15
 HOR3_HORVU STANDARD; PRT; 264 AA.
 AC P06471;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 83-hordein (Fragment).
 OS Hordeum vulgare (Barley).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 CC Triticaceae; Hordeum.
 CC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:44:16 ; Search time 44 Seconds
(without alignments)
345,210 Million cell updates/sec

Title: US-09-743-533-19
Perfect score: 906
Sequence: 1 MROLNPGSOELQSPQSYLQ.....PPQQPQPVPOQASCMSMV 158

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	830.5	91.7	260 2	seed storage prote
2	732	83.0	310 2	C-hordein precursor
3	580	64.0	347 2	probable hordein C
4	566	62.5	357 2	omega secalin prec
5	562.5	62.1	357 2	omega secalin prec
6	412	45.5	327 2	gamma-gliadin prec
7	384.5	42.4	302 2	gamma-gliadin prec
8	347	38.3	251 2	gamma-gliadin prec
9	332	36.6	291 1	EBMTG
10	310.5	34.3	105 2	gamma-gliadin B pr
11	302	33.3	305 2	C-hordein (clone p
12	289.5	32.0	307 2	alpha-hordein 1 pr
13	279.5	30.8	509 2	alpha/beta-gliadin
14	272.5	30.1	72 2	egg envelope prote
15	271.5	30.0	72 2	C-hordein (pc horl
16	268.5	29.6	293 2	hordein B1 - barle
17	263.5	29.1	326 2	alpha/beta-gliadin
18	262.5	29.0	297 2	alpha/beta-gliadin
19	260.5	28.8	320 2	hordein B precursor
20	256.5	28.3	286 2	alpha/beta-gliadin
21	255.5	28.2	319 2	gamma/beta-gliadin
22	255	28.1	296 2	alpha/beta-gliadin
23	253.5	28.0	296 2	gliadin - wheat
24	250	27.6	291 2	alpha/beta-gliadin
25	244.5	27.0	1343 2	cell division prot
26	242.5	26.8	374 2	glutenin low molec
27	242	26.7	282 2	glutenin low molec
28	240	26.5	282 2	B3-hordein (clone
29	239	26.4	313 2	alpha/beta-gliadin

30	238	26.3	356 2	S01992	glutenin low molec
31	238	26.3	359 2	T06982	glutenin low molec
32	237	26.2	288 2	T06282	alpha-gliadin prec
33	236.5	26.1	319 2	A22364	alpha/beta-gliadin
34	236	26.0	286 2	S07923	alpha/beta-gliadin
35	235.5	26.0	276 2	S57656	glutenin low molec
36	234	25.8	286 1	EBMTA	alpha/beta-gliadin
37	234	25.8	292 2	B22364	alpha/beta-gliadin
38	234	25.8	1342 2	G90750	cell division prot
39	234	25.8	1342 2	E85614	cell division prot
40	230.5	25.4	261 2	S57655	glutenin low molec
41	219	24.2	271 2	T04474	B1 hordein - barle
42	218.5	24.1	68 2	B25677	C-hordein (pc-919)
43	217	24.0	1329 2	A64828	cell division prot
44	211.5	23.3	307 2	S04325	glutenin low molec
45	208.5	23.0	462 1	LUD07	annexin VII - silm

ALIGNMENTS

RESULT 1
S18350
seed storage protein - barley
C:Species: Hordeum vulgare (barley)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C:Accession: S18350
R:Entwistle, J.; Knudsen, S.; Mueller, M.; Cameron-Mills, V.
Plant Mol. Biol. 17, 1217-1231, 1991
R>Title: Amber codon suppression: The in vivo and in vitro analysis of two C-hordein
A:Reference number: S18350; MUID:92032786; PMID:1932695
A:Accession: S18350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <SENT>
A:Cross-references: EMBL:X60037; NID:919000; PIDN:CAA42642.1; PID:q19001
C:Superfamily: gliadin

Query Match 91.7% Score 830.5; DB 2; Length 260;
Best local similarity 98.7% Pred. No. 7.7e-53;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 ROLNPGSOELQSPQSYLQDPYLPQKPPVQOPFHTPPYLPDELFPQYQIP 61
DB 21 ROLNPGSOELQSPQSYLQDPYLPQKPPVQOPFHTPPYLPDELFPQYQIP 80
QY 62 TPILQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 121
DB 81 TPILQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 140
QY 122 QIIFQPPQSPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 150
DB 141 QIIFQPPQSPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 168

RESULT 2
T06211
C-hordein precursor - barley
C:Species: Hordeum vulgare (barley)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C:Accession: T06211; A03355
R:Salnova, O.V.; Mekhedov, S.L.; Zhelnin, L.G.; Khokhlova, T.A.; Anan'ev, E.V.
Genetika 29, 1070-1079, 1993
A>Title: Nucleotide sequence of the barley C-hordein gene.
A:Reference number: Z15537; MUID:93380629; PMID:8396543
A:Accession: T06211
A:Status: preliminary; translated from GR/EMBL/DDB
A:Molecule type: DNA
A:Residues: 1-310 <SENT>
A:Cross-references: EMBL:S66938; PIDN:AA28161.1; PID:q4442524
R:Shewry, P.R.; Autran, J.C.; Nimmo, C.C.; Tew, E.C.L.; Kasarda, D.D.
Nature 286, 520-522, 1980
A:Reference number: A93228


```

JS0402
gamma-gliadin precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Nov-1996
C:Accession: JS0402
R:Sugiyama, T.; Katsuki, A.; Soell, D.
Plant Sci. 44, 205-209, 1986
A:Title: The nucleotide sequence of a wheat gamma-gliadin genomic clone.
A:Reference number: JS0402
A:Accession: JS0402
A:Molecule type: DNA
A:Residues: 1-327 <SUG>
A:Experimental source: cv. Yamhill
C:Comment: Gamma-gliadins belong to a subclass of the storage proteins.
C:Superfamily: gliadin
C:Keywords: storage protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-327/Product:gamma-gliadin #status predicted <MAT>

```

Query Match	45.5%	Score 412;	DB 2;	Length 327;
Best Local Similarity	50.0%	Pred. No. 7,4e-23;		
Matches	97;	Conservative 12;	Mismatches 37;	Indels 48;
				Gaps 10;

QY	3	QLNPSQSE--LQSPQASTLOQPYPONPYLPQKPPFVQOQPFHNPQOQFPYLPBELFPYOI	60
Db	22	QVDSGGQGWLQOQLVLPLOQLPSGQP-----QQTFPQPOQTFFHQPOQGVPPQOQ	72
QY	61	PFP--LQVQGFPPQOQPOQPLP---RPQGFPPQPOQPF---QPOGPIFQOQOQGFPP--Q	110
Db	73	PQOQPIQLQOQOQFPPOQPOQPFPPQOQOQPOQPFQOQOQPOQOQOQPFQOQPOQPFPPQOQ	132
QY	111	QPOQPFPP--GQPOQOIIIFQOPOQOS--YVPVQPOQPFPPQOQPPVFPQOQ-----	151
Db	133	QPOQGFPPQLQOQPOQ--PFPQPOQOQLPQPOQPOQOOSFPQOQRPQIFPSLQOQLNFCNKILLQO	191
QY	152	-----SCIMSMV	158
Db	192	SKPASLVSSLWSTII	205

```

RESULT 7
JA0153
gamma-gliadin precursor - wheat
N:Alternate names: seed storage protein
C:Species: Triticum aestivum (common wheat)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 15-Nov-1996
C:Accession: JA0153
R:Schneets, K.; Hedgcock, C.
Plant Sci. 57, 141-150, 1988
A:Title: Nucleotide sequence of a gamma gliadin gene: comparisons with other gamma gliad
A:Reference number: JA0153
A:Accession: JA0153
A:Molecule type: DNA
A:Residues: 1-302 <SCCH>
A:Experimental source: seed
C:Comment: Wheat storage proteins are classified as prolamins (alpha-/beta-, gamma- and
lity, molecular weight, and amino acid content.
C:Superfamily: gliadin
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-302/Product: gamma-gliadin #status predicted <GAG>

```

	Query Match	42.4%	Score 384.5;	DB 2;	Length 302;
	Best Local Similarity	58.0%;	Pred. No. 6.3e-21;		
	Matches 91; Conservative	11;	Mismatches 28;	Indels 27;	Gaps 11;
OY	3 QLNPCSGELOSPPQOSTLYLQAPYPQNNYLRLPKRPVQQPFT--PQGYPRYLPEELFPQYQ 59				
	:: ::	:: ::	:: ::	:: ::	
Dd	22 QVDP-SGVQWMPQ-----QQPFRQ-----QQPFR-CQOPORTIPRPNQTGTHNQOQQFFRPOFO 71				
	:: ::	:: ::	:: ::	:: ::	
OY	60 IPTPLQPPQPPP--QQPQPQLPRLRPQPPRPPQDQGF--QRQET--PQDQPPRP--QQP 112				
	:: ::	:: ::	:: ::	:: ::	
Dd	72 QTTHQPPQQAQPFQTIQQQPPRPPQPPQQTTPQQAQLPFRPPQPPRPPQPPQPPQPPQPSQSP 131				
	:: ::	:: ::	:: ::	:: ::	

```
QY 113 QDEFPQQPQQI1FQQPQOSIPVQQPQQPFPQOPQVPQ 149
      ||||| || |||| |||| ||||| |||
Db 132 QQPFPPQQQ--FPQPQ-----QPQQSFPQQQQPAIQ 161
```

```

RESULT 8
PS0094
gamma-gliadin precursor (clone pm10) - wheat (fragment)
C:Species: Triticum aestivum (common wheat)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 20-Aug-1999
C:Accession: PS0094
R:Schneets, K.; Rafalski, J.A.; Hedgcock, C.; Soell, D.G.
Plant Sci. Lett. 37, 221-225, 1985
A:Title: Heptapeptide repeat structure of a wheat gamma-gliadin.
A:Reference number: PS0094
A:Accession: PS0094
A:Molecule type: mRNA
A:Residues: 1-251 <SCH>
A:Cross-references: GB:M6060; NID:g170735; PIDN:AAA34288.1; PID:g170736
C:Comment: Gliadins are the major storage proteins of wheat grain.
C:Superfamily: gliadin
C:Keywords: storage protein
E:1-19/Domain: signal sequence #status predicted <SIG>
E:20-251/Product: gamma-gliadin (clone pm10) (fragment) #status predicted <MAT>

```

Query Match	38.3%	Score 347	DB 2	Length 251
Best Local Similarity	45.0%	Pred. No. 2,5e-18		
Matches	86	Conservative	11	Mismatches 24
			Indels	70
			Gaps	11

```

RESULT 9
EWMTG
gamma-gliadin B precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A25632
R:Rafalski, J.A.
Gene 43, 221-229, 1986
A:Title: Structure of wheat gamma-gliadin genes.
A:Reference number: A25632; MUID:86501876; PMID:3017812
A:Accession: A25632
A:Molecule type: DNA
A:Residues: 1-291 <RAF>
A:Cross-references: GB:ML3713; NID:q170707; PIDN:AAA34274.1; PID:q170708
C:Comment: Gliadin, a glutamine and proline rich protein, is a major storage protein.
C:Superfamily: gliadin
C:Keywords: seed; storage protein
E:1-19/Domain: signal sequence #status predicted <STG>
F:20-291/Product: gamma-gliadin B #status predicted <MAT>

```

[illegible]

C:Genetics:
 A:introns: 224/1: 285/3: 318/2: 372/1: 436/2: 485/2
 C:Superfamily: trefoll homology; 2P domain homology
 F:186-221/Domain: trefoll homology <TRES>
 F:227-505/Domain: 2P domain homology <2PH>

Query Match 30.8%; Score 279.5; DB 2: Length 509;
 Best Local Similarity 51.0%; Pred. No. 3.3e-13;
 Matches 76; Conservative 11; Mismatches 25; Indels 37; Gaps 14;

QY 16 QSTYLPQ-----YPONPYLPQKFFVQGPHTPQQTFFPYLPDELFPQYQIPTPLDPOQ 68
 DB 57 QRYVQQRRLRLHDSFQ-----PGAPPDTPQ--PTYPQ--PQ---QPODPOQ 100
 QY 69 P-FPQDPQPLPFPQDP-FPWQDPQFPFPQDP-TPQDPQDPFPQDPQDP-FPQDPQDII 124
 DB 101 PKYPQDPQDP-QDPQDPKYPQDPQDP-QDPQDPKYPQDPQDP-QDPQDPKYPQDP----- 152
 QY 125 PDPQDPQSTPVPQDPQDPQDP-QDPVDPQAS 152
 DB 153 -QDPKNPQKPNPQDPQKPNPQKQQVS 180

RESULT 14

A25677

C-hordein (pc hor1-3) - barley (fragment)

C:Species: Hordeum vulgare (barley)

C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 12-Apr-1995

C:Accession: A25677

R:Rasmussen, S. K.; Brandt, A.

Carlsberg Res. Commun. 51, 371-379, 1986

A:Title: Nucleotide sequences of cDNA clones for C-hordein polypeptides.

A:Reference number: A90768

A:Accession: A25677

A:Molecule type: mRNA

A:Residues: 1-72 <RAS>

C:Superfamily: gliadin

Query Match

30.1%; Score 272.5; DB 2: Length 72;

Best Local Similarity 65.4%; Pred. No. 1.6e-13;

Matches 53; Conservative 7; Mismatches 10; Indels 11; Gaps 3;

QY 79 PRPQDPFPMQDPQDPFPP-QDPQDPFPMQDPQDPFPPQDPQDPQDPQDPQDPQDPQDPQDP 137
 DB 2 PQDPQDPFPMQDPQDPFPPQDPQDPFPMQDPQDPFPPQDPQDPQDPQDPQDPQDPQDP 60
 QY 138 QPFPQDPQDPFPMQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 158
 DB 61 QPYTQD-----TIWSMV 72

RESULT 15

S07365

hordein B1 - barley

C:Species: Hordeum vulgare (barley)

C:Date: 02-Dec-1993 #sequence_revision 01-Dec-1995 #text_change 20-Aug-1999

C:Accession: S07365

R:Forde, B. G.; Heyworth, A.; Pywell, J.; Kreis, M.

Nucleic Acids Res. 13, 7327-7339, 1985

A:Title: Nucleotide sequence of a B1 hordein gene and the identification of possible ups

A:Reference number: S07365; MUID:86041918; PMID:4059057

A:Accession: S07365

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-293 <FOR>

A:Cross-references: EMBL:X03103; NID:q18909; PIDN:CAA26889.1; PID:q18910

C:Superfamily: gliadin

C:Keywords: seed; storage protein

Query Match

30.0%; Score 271.5; DB 2: Length 293;

Best Local Similarity 45.3%; Pred. No. 7.2e-13;

Matches 77; Conservative 13; Mismatches 25; Indels 55; Gaps 14;

QY 20 QDPVDPNPYLPQKDPFVQDPFHTPQQTFFPYLPDELFPQYQIPTPLDPOQDPQDP--PQDP 77
 DB 21 QDPFQDP--TPQDP-----QPY--PQDPQPY-PQDPF-----PQDPFQDPVDPQDP 63
 QY 78 LPRPQDPFPMQDPQDPFPP-QDPQDPFPMQDPQDPFPPQDPQDPQDPQDPQDPQDPQDPQDP 125
 DB 64 QPYPQDPFPP--PQDPFPMQDPFPMQDPFPMQDPFPMQDPFPMQDPFPMQDPFPMQDPFPM 121
 QY 126 Q-----QPOQSYF-----YOPQDPFPMQDP--PQDPF-----QDASC 153
 DB 122 QTLQDLQIOTYVHPSTLQQLNPKKVFLOQDCSPVPVDPQRIARBSQMLQDQSC 171

Search completed: May 27, 2003, 14:47:46
 Job time : 53 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:41:41 : Search time 31 Seconds
(without alignments)
1050.176 Million cell updates/sec

Title: us-09-743-533-19
Perfect score: 906
Sequence: 1 MKOLNPGSOELQSPQSYLQ.....PFPQPPQPVQOASCIWSMV 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:.*
2: SP_BACTERIA:.*
3: SP_FUNGI:.*
4: SP_HUMAN:.*
5: SP_INVERTEBRATE:.*
6: SP_MAMMAL:.*
7: SP_MHC:.*
8: SP_ORGANELLE:.*
9: SP_PHAGE:.*
10: SP_PLANT:.*
11: SP_PROTOZOA:.*
12: SP_VIRUS:.*
13: SP_VERTEBRATE:.*
14: SP_UNCLASSIFIED:.*
15: SP_VIRUS:.*
16: SP_BACTERIAP:.*
17: SP_ARCHAEP:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	830.5	91.7	260	10 Q40053	Q40053 hordeum vul
2	752	83.0	310	10 Q41210	Q41210 hordeum vul
3	580	64.0	347	10 Q40055	Q40055 hordeum vul
4	574	63.4	280	10 Q9FWU7	Q9FWU7 triticum ae
5	568.5	62.7	357	10 Q04655	Q04655 secale cere
6	566	62.5	357	10 Q05573	Q05573 secale cere
7	562.5	62.1	357	10 Q05573	Q05573 secale cere
8	457	50.4	455	10 Q9FR41	Q9FR41 secale cere
9	440	48.6	337	10 Q94G98	Q94G98 triticum ae
10	440	48.6	337	10 Q94G96	Q94G96 triticum ae
11	412	45.5	308	10 Q9MG7	Q9MG7 triticum ae
12	412	45.5	327	10 Q94G91	Q94G91 triticum ae
13	410.5	45.3	255	10 Q9FTC2	Q9FTC2 aegilops sp
14	393.5	43.4	256	10 Q9FTC1	Q9FTC1 aegilops sp
15	384.5	42.4	302	10 Q94G93	Q94G93 triticum ae
16	383.5	42.3	300	10 Q9FA8	Q9FA8 aegilops bi

17	383.5	42.3	300	10 Q9FWU1	Q9FWU1 aegilops 10
18	378.5	41.8	280	10 Q41602	Q41602 triticum tu
19	375	41.4	234	10 Q9FTC3	Q9FTC3 aegilops sh
20	361	39.8	229	10 Q9FTC4	Q9FTC4 aegilops se
21	361	39.8	231	10 Q9FTC5	Q9FTC5 aegilops se
22	357	39.4	285	10 Q94G94	Q94G94 aegilops ae
23	351	38.7	298	10 Q94G92	Q94G92 triticum ae
24	350	38.6	294	10 Q9FWU0	Q9FWU0 aegilops 10
25	347	38.3	243	10 Q9FTC0	Q9FTC0 aegilops la
26	347	38.3	259	10 Q9FS62	Q9FS62 triticum ae
27	347	38.3	264	10 Q9FS74	Q9FS74 triticum ae
28	347	38.3	267	10 Q9FEG9	Q9FEG9 triticum ae
29	347	38.3	267	10 Q9FS60	Q9FS60 triticum ae
30	347	38.3	270	10 Q9FS69	Q9FS69 triticum ae
31	347	38.3	283	10 Q9FS73	Q9FS73 triticum ae
32	347	38.3	295	10 Q9FEM0	Q9FEM0 triticum ae
33	347	38.3	298	10 Q9SYX8	Q9SYX8 triticum ae
34	343	37.9	279	10 Q41543	Q41543 triticum ae
35	341.5	37.7	259	10 Q9FS75	Q9FS75 triticum ae
36	338	37.3	243	10 Q9FMB9	Q9FMB9 aegilops la
37	337.5	37.3	252	10 Q9FS77	Q9FS77 triticum ae
38	335	37.0	192	10 Q9FS58	Q9FS58 triticum ae
39	335	37.0	250	10 Q9FS76	Q9FS76 triticum ae
40	335	37.0	254	10 Q9FS56	Q9FS56 triticum ae
41	334	36.9	244	10 Q9FS61	Q9FS61 triticum ae
42	332.5	36.7	248	10 Q9FS57	Q9FS57 triticum ae
43	332	36.6	192	10 Q9FS59	Q9FS59 triticum ae
44	324	35.8	563	13 Q919M8	Q919M8 oncorhynch
45	323.5	35.7	277	10 Q9M415	Q9M415 triticum ae

ALIGNMENTS

RESULT 1
ID Q40053 PRELIMINARY: PRT: 260 AA.
AC Q40053;
DT 01-NOV-1996 (TREMBLER, 01, Created)
DT 01-NOV-1996 (TREMBLER, 01, Last sequence update)
DT 01-DEC-2001 (TREMBLER, 19, Last annotation update)
DE Hor-17 C-hordein.
OS Hordeum vulgare (Barley).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Pooideae:
OC Triticeae: Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BOMT;
RX MEDLINE=94035184; PubMed=8220485;
RA "Muller M., Knudsen S.;
RT "The nitrogen response of a barley C-hordein promoter is controlled by
RT positive and negative regulation of the GCN4 and endosperm box.";
RL Plant J. 4:343-355(1993).
DR EMBL: X60037; CAA42642.1; -;
SQ
SEQUENCE 260 AA; 30397 MW; CB220BD2309280AE CRC64;

Query Match 91.7%; Score 830.5; DB 10; Length 260;
Best Local Similarity 98.7%; Pred. No. 4.1e-69;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 KOLNPGSOELQSPQSYLQPPQNYLPPQKPPVVOQPHHTPPQYPPYLPPELFPQYQIP 61
DB 21 KOLNPGSOELQSPQSYLQPPQNYLPPQKPPVVOQPHHTPPQYPPYLPPELFPQYQIP 80

QY	62	TTTTPOQPFPPQOQOPLPRPQOQPFPMQOQOFPQPOQETIPQOPOQPFQOQOQPFPPQOQO	121
Db	81	TTTTPOQOQPFPPQOQOPLPRPQOQPFPMQOQOQPFQPOQETIPQOQOQPFPPQOQOQPFPPQOQO	140
OY	122	QITFQOQOQSTPVQPOQOQPFPPQOQOQPOQVPPQO	150
Db	141	QITFQOQOQSTPVQPOQOQPFPPQOQOQPOQVPPQO	168
RESULT 2			
ID	Q41210	PRELIMINARY;	PRT; 310 AA.
AC	Q41210;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	C-hordein.		
GN	C-HORDEIN.		
OS	Hordeum vulgare (Barley).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;		
OC	Triticeae; Hordeum.		
OX	NCBI_Taxid=4513;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93380629; PubMed=8396543;		
RA	Sainova O.V., Mekhedov S.L., Zheinin L.G., Khokhlova T.A.,		
RA	Anan'ev E.V.;		
RT	"Nucleotide sequence of the barley C-hordein gene".;		
RL	Genetika 29:1070-1079(1993).		
DR	EMBL: S66938; AAB28161.1;		
SO	SEQUENCE 310 AA; 36508 MW; 7A9621869D50D672 CRC64;		

Query Match	Similarity	83.0%	Score 752	DB 10	Length 310
Best Local	143	Conservative	2	Mismatches	3
				Indels	96
				Gaps	2

```

RP      SEQUENCE FROM N.A.
RC STRAIN-BOMI : TISSUE=IMMATURE ENDOSPERM;
RX MEDLINE=69351278; Pubmed=3255313;
RA Entwistle J.;
RT "Primary structure of a C-hordein gene from barley."
RL Carlisberg Res. Commun. 53:247-256(1988).
DR EMBL; M36941; AAA9233.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN          1         20       POTENTIAL.
FT              21        347       C HORDEIN.
SO SEQUENCE      347 AA; 40546 MW; 1E48919B2BCBD9D CNC64;

Query Match
Best Local Similarity    64.0%; Score 580; DB 10; Length 347;
Matches 119; Conservative 10; Mismatches 20; Indels 70; Gaps 6;

QY      2 ROLNCSOELDSP-----QQSYLDQPYRPNMYLKDFPRVQQRPHPTGQRYLPPELF 55
DB      21 ROLNISHOELSPQDPFLKQOSTYLQDPFRPQDYLRQQPRF-----TFQGFPRYLPQDTF 74
QY      56 POLYQIPTRPLQPOQPF---PQOPQDLPRKPQOPFFMQPQOPQPF-DQEPIRQQPQOPFPQO 111
DB      75 PRSQGPENPLQPGQPPRLQDPFRPQDGQGFPRQGNPNQDPQDPFRFGQQLIVQDQDPFPQO 134
QY      112 PQQPRP-----Q 118
DB      135 PQQPRPQOPPSWQPOQPFLLQPLQLXPLQAQAPFRLQDPQLPRPOPQDPISGOQPKQPLIQ 194
QY      119 QPQGLIFQDPQDSYFVPQOPQPFQOP-QPVPOQASITWS 156
DB      195 QPQQTITPQDPQOPFLQDPQOPFPQDPQOPQDLRQDPQILIS 233

```

[illegible]

```

ID      004365      PRELIMINARY:      PRT:      357 AA.
AC      004365:
DT      01-JUL-1997 (Tremblrel. 04, Created)
DT      01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT      01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE      Omega secalin.
OS      Secale cereale (Rye).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC      Triticaceae; Secale.
OX      NCBI_TaxID=4550;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-ENDOSPERM;
RX      MEDLINE=97094352; PubMed=8939819;
RA      Clarke B.C., Mikal Y., Appels R.;
RT      "The Sec-1 locus on the short arm of chromosome 1R of rye (Secale
RT      cereale).";
RL      Chromosoma 105:269-275(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE-ENDOSPERM;
RA      Clarke B., Appels R.;
RT      "Sequence variation at the Sec-1 locus of rye.";
RL      Plant Syst. Evol. 0:0-0(1997).
DR      EMBL: AF000227; AAB58403.1; -.
SQ      SEQUENCE      357 AA; 41625 MW;  E3219653C105EF66 CRC64;

```

	Overy Match:	62.7%	Score	568.5;	Df.	10;	Length	357,				
	Best Local Similarity	68.6%;	Pred.	No. 7.le-45;								
	Matches 109; Conservative	11;	Mismatches	30;	Indels	9;	Gaps	3				
OY	2 ROLNPSQELQSFGSYLQQPYPNWLYLRKPFVQPDPFHTPPOYPFYLPBELFPYOYL P : : : :											
Dd	20 ROLNPSQEQLSP-----DQPVKKESGYRPPRPESHGFPRTPROOYSYOPQOPFOGP 74											
OY	62 TPLOPQOPFFQOPGQPLRLRRQRFRMWRQORPORPORPEFTIRPOOOOPRPOGR--FFQ 118											
Dd	75 TPILOPQOPFFQOPGQPLRFSSOPDOOLRLQDRQOPPRPDPTIPDOOPDSFPQOPQOPREODQFPQ 134											
OY	119 QPOPILFPQOPQOSYVPQOPRFPQPGQP-PVDAQASCITWS 156											
Dd	135 QPOPILFPQOTQOFFPLQOPFFPLQOPRFAAQOREIIS 173											
<hr/>												
RESULT 6												
ID	005573	PRELIMINARY:	PRT.	357 AA.								
AC	005573;											
DT	01-NOV-1996 (TREMBLrel. 01, Created)											
DT	01-NOV-1996 (1-NOV-1996 (Last sequence update))											
DE	01-DIC-2001 (TREMBLrel. 19, Last annotation update)											
DD	Scc1 precursor.											
GN	SRC1											
OS	Secale cereale (rye).											
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;											
CC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;											
OC	Triticaceae; Secale.											
OX	NCBI_TaxID=4550;											
RN	[1]											
RP	SEQUENCE FROM N.A.											
RC	SRRAIN-CV, GAZELLE; TISSUP-ENDOSPERRA;											
RX	MEDLINE=92032773; PubMed=1932685;											
RA	Hall G.A., Halford N.G.; Kreis M., Shewry P.R. ;											
KT	"Isolation and characterisation of genes encoding rye prolamins											
RT	containing a highly repetitive sequence motif." ;											
RL	Plant Mol. Biol. 17:1111-1115(1991).											
KM	EMBL: X60295; CAA42837.1; -											
DR	SIGNAL.											
FT	CHAIN	1	19	POTENTIAL.								
SO	SEQUENCE	357 AA;	41439 MW;	BEC406DA8F72909E CRC64;								

	Query Match	Similarity	62.5%	Score 566;	DB 10;	Length 357;
	Best Local	Similarity	70.7%;	Pred No.	1.2e-44;	
	Matches 104;	Conservative	11;	Mismatches 24;	Indels 8;	Gaps 2.
QY	2	ROLNPCSOELQSPQOOSYLQAPYPONFVLPKQKFFPVYOQPFHTFTQOXYFPYLPBELFPQYQIP	61			
Db	20	ROLNPSEQLGDP-----QQPYPKBGSYRQOQYPSHQPPRFPOQYSFYVQDQDFPQPDP	74			
QY	62	TPLQDPQPPFPQOPQDLPRPQDPFFWQPOQPPFPQDEPIPLQODQDPFFQOPQOP---FPQ	118			
Db	75	APIQPQPPFPQOPQOPFFPQPOOOLPLQDPQPPFPQOPQIPLQDPQOSFFPQOPQOPRPEQOPQ	134			
QY	119	OPQGIIFPQPOQSYPVQOPQPPFPQOPQ	145			
Db	135	QPQGIILPQOTQOPFLQDPQPPFPQOPQ	161			

RESULT 7		
ID	043639	PRELIMINARY: PRT: 357 AA.
AC	043639;	
DT	01-NOV-1996 (TREMBlrel. 01, Created)	
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)	
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)	
DE	Secl precursor.	
GN	SECL.	
OS	Secale cereale (Rye).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spemmatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;	
OC	Trilicaceae; Secale.	
OX	NCBI_TaxID=4550;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. GAZELLE; TISSUE=ENDOSPERM;	
RA	MEDLINE=92032773; PubMed=1932685;	
RX	Hull G.A., Halford N.G., Kreis M., Shewry P.R.;	
RT	"Isolation and characterisation of genes encoding rye prolamins	
RL	containing a highly repetitive sequence motif.";	
DR	Plant Mol. Biol. 17:111-115(1991).	
DR	EMBL: X60294; CAA42836.1; -;	
KW	Signal.	
FT	SIGNAL.	1 19
FT	CHAIN	20 357
FT		POTENTIAL.
FT		POTENTIAL.
SEQ	SEQUENCE	357 AA: 41541 MW: 795705SCAI005BIE CRC64:

Query Match	62.1%	Score 562.5	DB 10	Length 357
Best Local Similarity	67.9%	Pred. No. 2.5e-44		
Matches 108	Conservative 11	Mismatches 31	Indels 9	Gaps 3
OY	2	ROLNCCSGELOSRRQSSYLQRRYRONPULRYKRRFRYQRRPNTROXYRPRRLPRYOYIR	61	
		: : :		
Db	20	ROLNNSBQFLQSP-----OQPVKKEQSTPRQOYRPSHQRPRRQOYSFYQRRQRRPQRR	74	
OY	62	TRLQRRPFRFRQRRQRLPRFRQRRPFRFRQRRPFRFRQRRPFRFRQRRPFRFRQRRPFR	118	
		:		
Db	75	TRLQRRPFRFRQRRQRRPFRFRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRR	134	
OY	119	QRPQIIFQRRQSSYRVRQRRQRRFRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRR	156	
		: :		
Db	135	QRPQIIFQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRRQRR	173	
RESULT 8				
ID	O9FR41	PRELIMINARY:	PRT:	455 AA.
OC	O9FR41			
DT	01-MAR-2001	(TREMBLrel, 16, Created)		
DT	01-MAR-2001	(TREMBLrel, 16, Last sequence update)		
DT	01-MAR-2002	(TREMBLrel, 20, Last annotation update)		
DE	Secalin.			
GN	GSEC2-1.			
OS	Secale cereale (Rye).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Secale.
 OX NCBI_TaxID=4550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. IMPERIAL;
 RA Murray F.R., Skeritt J.H., Appels R.;
 RT "A gene from the Sec2 (Gli-R2) locus of a wheat 2RS.2BL chromosomal
 translocation line."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF201084; AAG35598.1; -
 DR InterPro: IPR001954; G1a.gluTenin.
 DR InterPro: IPR002965; P.rich.extensin.
 DR InterPro: IPR001768; Try/amyL_inhtr.
 DR Pfam: PF00234; tryp_alpha_amyL.1.
 DR PRINTS: PR00208; GLTADGLUTEN.
 DR PRINTS: PR01217; PRICHEXTENSIN.
 FT CHAIN 20 455
 SQ SEQUENCE 455 AA; 51731 MW; C1A176DD045D927 CRC64;

Query Match 50.4%; Score 457; DB 10; Length 455;
 Best Local Similarity 61.8%; Pred. No. 1,6e-34;

Matches 94; Conservative 12; Mismatches 34; Indels 12; Gaps 6;

OY 3 QLNPCSOELQSPQSYLQDPYQNPYLPQKPPVQOPFHTPQOYFPLYPEELFPOYOIPT 62
 DB 22 QVNP-SGQVQCCQOQPPQPOQSSSQOQPOQPPQOS-----QQPFPQOQOSSFPOQOP- 74
 OY 63 PLQOPQFPQOPQOPLP-RPQOPFPWQOPQPP-OPQEPQOQPOQPPQOQPOQFPQOP 120
 DB 75 --YPQOPFPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOP 132
 OY 121 QQIITQOQPOQSPVQOPQOPFPQOP-QPVPQQA 151
 DB 133 QQPFPQOPLQOPQOPQOPFPQOPQOPQOPQOS 164

RESULT 9

ID 094G98 PRELIMINARY; PRT; 241 AA.

AC 094G98;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Gamma-gliadin (Fragment).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D., Hsia C., Torres V.;
 RT "The wheat gamma-gliadin genes: characterization of ten new sequences
 RT and further understanding of gamma-gliadin gene family structure."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF234642; AAK84772.1; -
 FT NON_TER 241
 SQ SEQUENCE 241 AA; 28079 MW; A2765A48A35FD6D3 CRC64;

Query Match 48.6%; Score 440; DB 10; Length 241;
 Best Local Similarity 57.5%; Pred. No. 3.2e-33;

Matches 100; Conservative 12; Mismatches 26; Indels 36; Gaps 11;

OY 3 QLNPCSOELQSPQSYLQDPYQNPYLPQKPPVQOP---FHTPQOYFPLYPEELF----- 55
 DB 22 QVNP-SGQVQCCQOQPPQPOQSSSQOQPOQPPQOS-----QQPFPQOQOSSFPOQOP 72
 OY 56 -----PQYQITPPLQOPQOPFPQOPQOPLP---RPQOPFPWQOPQOPFPQOP-----PIP 101
 DB 73 QTFPHQPOQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOP 132

OY 102 QOPQOPFPQOPQOPF---PQOPQOI---FQOPQO---SYVPQOPQOPFPQOPQOPVQ 149
 DB 133 QOPQOPFPQOPQOPFPQOPQOPQOPFPQOPQOPQOPQOPQOPQOPQOPQOPQOP 186

RESULT 10

ID 094G96 PRELIMINARY; PRT; 337 AA.

AC 094G96;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Gamma-gliadin.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D., Hsia C., Torres V.;
 RT "The wheat gamma-gliadin genes: characterization of ten new sequences
 RT and further understanding of gamma-gliadin gene family structure."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF234644; AAK84774.1; -
 DR InterPro: IPR001768; Try/amyL_inhtr.
 DR Pfam: PF00234; tryp_alpha_amyL.1.
 SQ SEQUENCE 337 AA; 38634 MW; 524DCFA9CE1F0C13 CRC64;

Query Match 48.6%; Score 440; DB 10; Length 337;
 Best Local Similarity 57.5%; Pred. No. 4.4e-33;

Matches 100; Conservative 12; Mismatches 26; Indels 36; Gaps 11;

OY 3 QLNPCSOELQSPQSYLQDPYQNPYLPQKPPVQOP---FHTPQOYFPLYPEELF----- 55
 DB 22 QVNP-SGQVQCCQOQPPQPOQSSSQOQPOQPPQOS-----QQPFPQOQOSSFPOQOP 72
 OY 56 -----PQYQITPPLQOPQOPFPQOPQOPLP---RPQOPFPWQOPQOPFPQOP-----PIP 101
 DB 73 QTFPHQPOQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOP 132
 OY 102 QOPQOPFPQOPQOPF---PQOPQOI---FQOPQO---SYVPQOPQOPFPQOPQOPVQ 149
 DB 133 QOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOPFPQOPQOP 186

RESULT 11

ID 09M6P7 PRELIMINARY; PRT; 308 AA.

AC 09M6P7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Gamma-gliadin (Fragment).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YAMHILL;
 RA Arentz-Hansen E.H., McAdam S.N., Molberg O., Kristiansen C.,
 RA Sollid L.M.;
 RT "Production of a panel of recombinant gliadins for the
 RT characterization of T cell reactivity in coeliac disease."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF175312; AAF42989.1; -
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001954; G1a.gluTenin.
 DR InterPro: IPR001768; Try/amyL_inhtr.
 DR Pfam: PF00234; tryp_alpha_amyL.1.

FT NON_TER 256 256
 SQ SEQUENCE 256 AA; 30456 MW; 8D2B52ED7B6CF9E0 CRC64;

Query Match 43.4%; Score 393.5; DB 10; Length 256;
 Best Local Similarity 51.4%; Pred. No. 6,4e-29;
 Matches 94; Conservative 13; Mismatches 21; Indels 55; Gaps 13;

```
OY 12 QSPQSYL--QPPYPNPLRQKRPYQQRPHPPQYFPL--PEELFPQYQIPTPLPQ 67
    | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 32 QPPQQLFQPPQPPQ--PQPPYPO--PQPPPTQPPQLFQSQ-----QPQ 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 68 QPPQPPQPPPLP--RPPQFP--WPPQFP--QPPPIPPQPPQPPFP--QPPQPPPPQ 119
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 79 QPPQPPQPPYPPQPPQPPFPPTQPPQLFPQSQPPQPPYPPQPPQPPPPQPPQPPQPP 137
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 120 PQQIIFPPQPPQSY--PVPQPPPPQPP--QPPVQ-----ASCIW 155
    : | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db 138 -----YQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 192
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 156 SMV 158
    | | |
Db 193 SMT 195
```

RESULT 15

```
O94G93 PRELIMINARY; PRT: 302 AA.
ID O94G93
AC O94G93;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Gamma-gliadin.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHEYENNE;
RA Anderson O.D., Hsia C., Torres V.;
RT "The wheat gamma-gliadin genes: characterization of ten new sequences
RT and further understanding of gamma-gliadin gene family structure.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF234647; AAK84777.1; -.
DR InterPro: IPR001768; Try/amyL_inhblt.
DR Pfam: PF00234; Tryp_alpha_amiL; 1.
SQ SEQUENCE 302 AA; 34286 MW; 663B37C768BAE782 CRC64;
```

Query Match 42.4%; Score 384.5; DB 10; Length 302;

Best Local Similarity 58.0%; Pred. No. 5.1e-28;

Matches 91; Conservative 11; Mismatches 28; Indels 27; Gaps 11;

```
OY 3 QLNPCSQELQSPQSYLQPPYPPYLPQKPPPVQDPFHT---PQYFPYLPBELFPQYQ 59
    | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db 22 QVDP--SGQVWMP--QPPFP--PQPP--COQPPQTIPOPHQTFHNPQPPFPQ 71
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 60 IPPPLPQPPFP--QPPQPPLPQPPQPPQPPQPPFP--OPEPI--PQPPQPPFP--QPP 112
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 72 QYVPHQPPQPPQPPQPPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 131
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 113 QQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 149
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 QQPPFPQPPQ--FPQPPQ-----QPPQSPQPPQPPQPPQPPQPPQPPQPP 161
```

Search completed: May 27, 2003, 14:46:48
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:47:50 ; Search time 518.477 Seconds

(without alignments)
7858.394 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
Sequence: 1 atagatacagcagatgcctccc.....gagctcgattgcctcata 140

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Genembi:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hum:*
- 31: em_hum:*
- 32: em_hum:*
- 33: em_hum:*
- 34: em_hum:*
- 35: em_hum:*
- 36: em_hum:*
- 37: em_hum:*
- 38: em_hum:*
- 39: em_hum:*
- 40: em_hum:*
- 41: em_hum:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.6	25.4	388	8	HVE1393VR
2	35.6	25.4	558	8	HVE1326PR
3	35.2	25.1	1222	6	AR170979
4	34.4	24.6	2050	1	AP180731
5	34.2	24.4	163	6	AR210350
6	33.8	24.1	484	6	AR147096
7	33.8	24.1	484	6	AR147096
8	33.4	23.9	1447	8	GCN132219
9	33.4	23.9	1447	8	GCN132212
10	33.4	23.9	5506	6	AR195508
11	33.4	23.9	5507	8	LSRGCD21
12	33.3	23.6	345	9	HSZ96972
13	32.6	23.3	163	6	AR210351
14	32.6	23.3	15394	2	AC107595
15	32.4	23.1	278	8	HV1371VRC
16	32.4	23.1	500	11	G74883
17	32.4	23.1	170241	2	AC016434
18	32.2	23.0	481	9	HSPPT06
19	32.2	23.0	700	6	AX392015
20	32.2	23.0	4350	12	AP173954
21	32.2	23.0	241901	2	AC128708
22	32.2	22.9	494	8	HVE13270R
23	32.2	22.9	541	9	CJSGP2
24	32.2	22.9	672	11	G68272
25	32.2	22.9	1432	3	HSC299386
26	32.2	22.9	3027	8	AP395756
27	32.2	22.9	3551	12	AR122058
28	32.2	22.9	153137	9	AC016995
29	31.8	22.7	1843	10	AP232061
30	31.4	22.4	490	8	CSA416715
31	31.4	22.4	608	1	UBA011367
32	31.4	22.4	7823	12	AF041426
33	31.4	22.4	8696	12	AF379854
34	31.4	22.4	40544	1	SCH5
35	31.2	22.3	1611	8	AF308658
36	31.2	22.3	215	6	AX156393
37	31.2	22.1	419	6	AR166312
38	31.2	22.1	420	6	AR166319
39	31.2	22.1	502	3	U4L427427
40	31.2	22.1	530	6	AX027636
41	31.2	22.1	597	6	AR148006
42	31.2	22.1	597	6	AX116848
43	31.2	22.1	624	8	OS040814
44	31.2	22.1	632	8	BNA489198
45	31.2	22.1	741	8	BNA489197

ALIGNMENTS

RESULT 1
LOCUS HVE1393VR
DEFINITION H. vulgare BARE-1 long terminal repeat DNA; ID E1333VR.
ACCESSION Z64569.1
VERSION Z64569.1 GI:2598547
KEYWORDS BARE-1 retrotransposon; long terminal repeat.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE
AUTHORS Suontemi, A., Schmidt, D., and Schulman, A.H.
TITLE BARE-1 insertion site preferences and evolutionary conservation of

RNA and cDNA processing sites
 JOURNAL Genetics 100 (1-3), 219-230 (1997)
 MEDLINE 98103449
 PUBMED 9440275
 REFERENCE 2 (bases 1 to 388)
 AUTHORS Schulman A.H.
 TITLE Direct Submission
 JOURNAL Submitted (24-JAN-1997) Schulman A.H., University of Helsinki, Institute of Biotechnology, P.O. Box 56, Biocentre 1, Viikinkaari 9, FIN-00014 Helsinki, Finland
 FEATURES
 source 1..388
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 LTR 142..>273
 /note="matches nt 1-131 of BARE-1 5' LTR, 5' terminus"
 repeat_region 142..273
 /transposon="BARE-1"
 BASE COUNT 102 a 94 c 87 g 105 t
 ORIGIN
 Query Match 25.4%; Score 35.6; DB 8; Length 388;
 Best Local Similarity 82.0%; Pred. No. 0.1;
 Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAAACC 60
 |||||||
 Db 22 GCATGCTCCCGCGCCATGCGCGGATGTCCTCGCCCACTACATC 71
 RESULT 2
 LOCUS HVE1326PR 558 bp DNA linear PLN 07-JUL-2002
 DEFINITION H.vulgare BARE-1 long terminal repeat DNA: ID E1326PR.
 ACCESSION Z84562
 VERSION Z84562.1 GI:2598543
 KEYWORDS BARE-1 retrotransposon: long terminal repeat.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 558)
 AUTHORS Suoniemi A., Schmidt D. and Schulman A.H.
 TITLE BARE-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites
 JOURNAL Genetics 100 (1-3), 219-230 (1997)
 MEDLINE 98103449
 PUBMED 9440275
 REFERENCE 2 (bases 1 to 558)
 AUTHORS Schulman A.H.
 TITLE Direct Submission
 JOURNAL Submitted (24-JAN-1997) Schulman A.H., University of Helsinki, Institute of Biotechnology, P.O. Box 56, Biocentre 1, Viikinkaari 9, FIN-00014 Helsinki, Finland
 FEATURES
 source 1..558
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 LTR <153..300
 /note="matches nt 1682-1829 of BARE-1 5' LTR, 3' terminus"
 repeat_region <153..300
 /transposon="BARE-1"
 misc_feature 417..548
 /note="flanking genomic sequence matches BARE-1a retrotransposon (Z17327) nt 308-439"
 BASE COUNT 139 a 133 c 129 g 157 t
 ORIGIN
 Query Match 25.4%; Score 35.6; DB 8; Length 558;
 Best Local Similarity 82.0%; Pred. No. 0.11;
 Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAAACC 60
 |||||||
 Db 110 GCATGCTCCCGCGCCATGCGCGGATGTCCTCGCCCACTACATC 159
 RESULT 3
 LOCUS AR170979/C 1222 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 83 from patent US 6297007.
 ACCESSION AR170979
 VERSION AR170979.1 GI:17909929
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1222)
 AUTHORS Waters, B., Miao, Y., Ho, Y., Mai, and Tong, S., Kah.
 TITLE Method for isolation of biosynthesis genes for bioactive molecules
 JOURNAL Patent: US 6297007-A 83 02-OCT-2001;
 FEATURES
 source 1..1222
 /organism="unknown"
 BASE COUNT 272 a 361 c 319 g 270 t
 ORIGIN
 Query Match 25.1%; Score 35.2; DB 6; Length 1222;
 Best Local Similarity 92.5%; Pred. No. 0.17;
 Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGG 50
 |||||||
 Db 1176 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGG 1137
 RESULT 4
 LOCUS AF180731 2050 bp DNA linear BCT 01-OCT-1999
 DEFINITION Klebsiella pneumoniae dihydrofolate reductase (dhfrXII) and streptomycin 3'-adenylyltransferase (aadA2) genes, complete cds;
 ACCESSION AF180731
 VERSION AF180731.1 GI:6003514
 KEYWORDS
 SOURCE Klebsiella pneumoniae.
 ORGANISM Klebsiella pneumoniae.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Klebsiella.
 REFERENCE 1 (bases 1 to 2050)
 AUTHORS Kim, J.
 TITLE Direct Submission
 JOURNAL Submitted (26-AUG-1999) Microbiology, Coll. Med., Dankook University, San29, Anseo-dong, Cheonan, Choongnam 330-180, Republic of Korea
 FEATURES
 source 1..2050
 /organism="Klebsiella pneumoniae"
 /strain="Kps15"
 /db_xref="taxon:573"
 /transposon="class I integron"
 misc_feature 128..131
 /note="putative integrase target site; GTPA-sequence"
 gene 137..634
 /gene="dhfrXII"
 CDS 137..634
 /gene="dhfrXII"
 /function="trimethoprim resistance"
 /codon_start=1
 /transl_table=11
 /product="dihydrofolate reductase"
 /protein_id="AA00479.1"
 /db_xref="GI:6003515"
 /translation="MNSSESVRIYIVAAANGARNVITONGNPIMKIPGEOKIFRRLTECK VVVGKRKTPEISGKPLPNRHVLIVISROANYRATGCVVSTLSHAIALASDEGNELYYA

```

GAELIYTLALPHANGVLESEVHQTFEGDAFFPMLNETEFELVSTETIQAVIPYTHSV
ARRNG"
stem_loop      648..679
misc_feature    712..715
                /note="putative integrase target site; GTTA-sequence"
CDS
                746..1036
                /note="orf"
                /codon_start=1
                /transl_table=1
                /product="unknown"
                /protein_id="AAF00480.1"
                /db_xref="GI:6003516"
                /translation="MEIQAFSGVIOCLPCLPSGLHGLRRFSVFLASSPCVASA
                1032..1035
                /note="putative integrase target site; GTTA-sequence"
                1044..1833
                /gene="aadA2"
                1044..1047
                /gene="aadA2"
                1054..1833
                /gene="aadA2"
                /note="aminoacyl-adenyl[yl]transferase A"
                /codon_start=1
                /transl_table=1
                /product="streptomycin 3'-adenyl[yl]transferase"
                /protein_id="AAF00481.1"
                /db_xref="GI:6003517"
                /translation="MTIEVSNQSEVLSVIERHLESTLLAVHLGSAVDGKPYSDI
                DLIVYAVKDETRRALNDMEASAFPGSEETLRA1EVLVYHDDITFWRPAKRE
                LQEGEWMNDLIGIFEPAMIDILAILLAKRHSAVALGPAEEFEDVPEDELE
                ALRETLKLNQSPAGDERNVVLTLSINWYATGKIAEPDVADCAIKRLPAOYOP
                VLLEAKQAVYLOKEDHLASRADHLEEFIRFVCEITKSVK"
                1836..1891
stem_loop      468 a 545 c 525 g 512 t
BASE COUNT
ORIGIN
Query Match      24.6%; Score 34.4; DB 1; Length 2050;
Best Local Similarity 97.2%; Pred. No. 0.38;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCAT 46
Db 1 GCATGCTCCGGCGCCATGCGCGGATTCAT 36

RESULT 5
LOCUS AR210350 163 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 51 from patent US 6387657.
ACCESSION AR210350
VERSION AR210350.1 GI:21512558
KEYWORDS
SOURCE
ORGANISM Unknown.
SOURCE
REFERENCE
1 (bases 1 to 163)
Bocstein,D.A., Cohen,R.L., Goddard,A.D., Gurney,A.L., Hillan,K.J.,
Lawrence,D.A., Levine,A.J., Pennica,D., Roy,M.Ath. and Wood,W.L.
TITLE WISP polypeptides and nucleic acids encoding same
JOURNAL Patent: US 6387657-A 51 14-MAY-2002;
FEATURES
source 1..163
Location/Qualifiers
BASE COUNT 35 a 45 c 46 g 37 t
ORIGIN
Query Match      24.4%; Score 34.2; DB 6; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.27;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCATGAG 49
Db 11 GCATGCTCCGGCGCCATGCGCGGATTCATGAG 49

```

```

Db 41 GCATGCTCCGGCGCCATGCGCGGATTCATCACTAG 79

RESULT 6
LOCUS AR147996 484 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 175 from patent US 6225054.
ACCESSION AR147996
VERSION AR147996.1 GI:15112086
KEYWORDS
SOURCE
ORGANISM Unknown.
SOURCE
REFERENCE
1 (bases 1 to 484)
Frudakis,T.N., Smith,J.M. and Reed,S.G.
TITLE Compositions and methods for the treatment and diagnosis of breast
cancer
JOURNAL Patent: US 6225054-A 175 01-MAY-2001;
FEATURES
source 1..484
Location/Qualifiers
BASE COUNT 101 a 138 c 126 g 118 t 1 others
ORIGIN
Query Match      24.1%; Score 33.8; DB 6; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.46;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCATGAGGCACTAAACCTTG 64
Db 25 GCATGCTCCGGCGCCATGCGCGGATTCATGAGGCACTAAACCTTG 78

RESULT 8
LOCUS GGN132219 1271 bp mRNA linear PLN 23-NOV-1999
DEFINITION Gnetum gnetum mRNA for putative MADS domain transcription factor
GNM13.
ACCESSION AJ132219

```

VERSION	AT132219.1	GI:5019463
KEYWORDS	floral homeotic gene; gnm13 gene; MADS domain transcription factor.	
SOURCE	Gnetum gnemon	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Gnepophyta; Gnepsida; Gnetales; Gnetaaceae; Gnetum.	
REFERENCE	1 (bases 1 to 1271)	
AUTHORS	Winter,K.U., Becker,A., Munster,T., Kim,J.T., Saedler,H. and Theissen,G.	
TITLE	MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (13), 7342-7347 (1999)	
MEDLINE	99307411	
PUBMED	10377416	
REFERENCE	2 (bases 1 to 1271)	
AUTHORS	Theissen,G.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-JAN-1999) Theissen G., Molecular Plant Genetics, Max-Planck-Institut für Breeding Research, Carl-von-Linne Weg 10, 50829 Koeln, GERMANY	
FEATURES	Location/Oualifiers	
SOURCE	1..1271 /organism="Gnetum gnemon" /sub_species="gnemon" /db_xref="taxon:3382" 167..880 /gene="gnm13" 167..880 /gene="gnm13" /note="floral homeotic gene" /codon_start=1 /product="putative MADS domain transcription factor GNM13" /protein_10="CAB4459.1" /db_xref="gi:5019464" /translation="MGKGLIEIKRIENTTNRQVYFSKRGGILKAHELSQLDAELGS LIIFSSSKLIEYSASSSMKKILERYGVGARITVEDNHLCEMT RKNENEEKLQ TNIRRMGEDITSITMTEIHLHGOOLSAASRVSRKNOMLOLENI.RRKERILLEPOO NSHCRLLAEDQAAVEGYOEPLFEGVFCPPDPDKTAANAAGLHGHLPARLPD TOPNDQESTVPNRPVLDT"	
BASE COUNT	395 a 271 c 297 g 308 t	
ORIGIN		
Query Match	23.9%; Score 33.4; DB 8; Length 1271;	
Best Local Similarity	97.1%; Pred. No. 0.78;	
Matches	34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Oy	11 GCATGCTCCCGGCCGCATGGCCGGGGATTGTCA 45 42 GCATGCTCCCGGCCGCATGGCCGGGGATTGCCA 76	
Db		
RESULT 9		
LOCUS	GNM132212	
DEFINITION	Gnetum gnemon mRNA for putative MADS domain transcription factor	
ACCESSION	UJ132212	
VERSION	AJ132212.2	
KEYWORDS	gnm6 gene; MADS domain transcription factor.	
SOURCE	Gnetum gnemon.	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Gnepophyta; Gnepsida; Gnetales; Gnetaaceae; Gnetum.	
REFERENCE	1 (bases 1 to 1447)	
AUTHORS	Winter,K.U., Becker,A., Munster,T., Kim,J.T., Saedler,H. and Theissen,G.	
TITLE	MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (13), 7342-7347 (1999)	
MEDLINE	99307411	
PUBMED	10377416	
REFERENCE	2 (bases 1 to 1447)	
AUTHORS	Theissen,G.	

TITLE	Direct Submission					
JOURNAL	Submitted (12-JAN-1999) Theissen G., Molecular Plant Genetics, Max-Planck-Institut für Breeding Research, Carl-von-Linne Weg 10, 50829 Koeln, GERMANY revised by (4) 3 (bases 1 to 1447)					
REMARK						
REFERENCE	Becker, A.					
AUTHORS	Direct Submission					
TITLE	Submitted (22-NOV-1999) Becker A., Molecular Plant Genetics, Max-Planck-Institut für Breeding Research, Carl-von-Linne Weg 10, 50829 Koeln, GERMANY					
JOURNAL	On Nov 23, 1999 this sequence version replaced gi:5019436. Location/Qualifiers					
COMMENT	1..1447					
FEATURES	/organism="Gnetum gnetum"					
source	/sub-species="gnemon"					
	/db_xref="taxon:3382"					
gene	306..1037					
CDS	/gene="gqm6"					
	306..1037					
	/gene="gqm6"					
	/codon_start=1					
	/product="putative MADS domain transcription factor GQM6"					
	/protein_id="CA84452.2"					
	/db_xref="GI:6468288"					
	/translation="MGKGLAMKITEKNSROYTFSRKNGLKKYVELSLIGAGFIA LVTSNGLKLIVSHVKHGSLSQIITHRYLONPHALRVDQIFDTLLTYAKEDKRFDLR FFSDRLMOLESVPNI.ELOSLDELDLATYKVRKEEAAAYDSLMDLELMRK ONSLOKELRRKELLEDRHALOMFOOLMVDPPGISISOMESSCVSYSDRTSTFL ALOVSFPAVSSFRSVKPKFTLLNL21"					
BASE COUNT	424 a	299 c	351 g	372 t	1 others	
ORIGIN						
Query Match	23.9%, Score 33.4; DB 8; Length 1447;					
Best Local Similarity	86.0%; Pred. No. 0.81;					
Matches	37; Conservative 0; Mismatches 6; Indels 0; Gaps 0;					
Qy	11 GCATGCTCCCGGGCCCATGGCCGGGATTGTGCATGAGCAA 53 205 GCATGCTCCCGGGCCCATGGCCGGGATTGTAAC TTGGCGA 247					
Dd						
RESULT 10						
LOCUS	AR195508/c 5506 bp DNA linear PAT 20-APR-2002					
DEFINITION	Sequence 93 from patent US 6350933.					
ACCESION	AR195508					
VERSION	AR195508.1 GI:20244945					
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1. (bases 1 to 5506)					
TITLE	Michelmore,R.W., Shen,K.A. and Meyers,B.C. RG polynucleotides for conferring powdery mildew resistance in plants					
JOURNAL	Patent: US 6350933-A 93 26-FEB-2002:					
FEATURES	Location/Qualifiers					
source	1..5506					
BASE COUNT	1797 a	940 c	1051 g	1717 t	1 others	
ORIGIN						
Query Match	23.9%; Score 33.4; DB 6; Length 5506;					
Best Local Similarity	97.1%; Pred. No. 1.1;					
Matches	34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Oy	11 GCATGCTCCCGGGCCCATGGCCGGGATTGTCA 45 Dd 5468 GCATGCTCCCGGGCCCATGGCCGGGATTGTAA 5434					
RESULT 11						

LSRGC2D1/C
LOCUS LSRGC2D1 5507 bp DNA linear PLN 12-JAN-1999
DEFINITION Lactuca sativa RGC2D pseudogene, partial sequence.
ACCESSION AF113953 AF072270
VERSION AF113953.1 GI:4150855
KEYWORDS
SEGMENT 1 of 2
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.
REFERENCE 1 (bases 1 to 5507)
AUTHORS Meyers,B.C., Chin,D.B., Shen,K.A., Sivaramakrishnan,S., Lavelle,D.O., Zhang,Z. and Michelmore,R.W.
TITLE The major resistance gene cluster in lettuce is highly duplicated and spans several megabases
JOURNAL Plant Cell 10 (11), 1817-1832 (1998)
MEDLINE 99030190
PUBMED 9811791
REFERENCE 2 (bases 1 to 5507)
AUTHORS Meyers,B.C., Shen,K.A., Rohani,P., Gaut,B.S. and Michelmore,R.W.
TITLE Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection
JOURNAL Plant Cell 10 (11), 1833-1846 (1998)
MEDLINE 99030191
PUBMED 9811792
REFERENCE 3 (bases 1 to 5507)
AUTHORS Meyers,B.V., Shen,K.A., Rohani,P.J., Baul,B. and Michelmore,R.W.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) Vegetable Crops, UC Davis, 148 Asmundson, Davis, CA 95616, USA
FEATURES
source
1..5507
/organism="Lactuca sativa"
/db_xref="taxon:4236"
/map="Dm3 resistance gene cluster"
BASE COUNT 1798 a 940 c 1051 g 1717 t 1 others
ORIGIN
Query Match 23.9%; Score 33.4; DB 8; Length 5507;
Best Local Similarity 97.1%; Pred. No. 1.1;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTGTCA 45
|||||
Db 5468 GCATGCTCCGGCGCCATGCGCGGATTGTAA 5434

RESULT 12
LOCUS HS296972 345 bp mRNA linear PRI 05-MAY-2001
DEFINITION H. sapiens mRNA for immunoglobulin heavy chain variable region (43-8; VH3, 3-15/DP-38).
ACCESSION Z96972.1 GI:2370221
VERSION Z96972.1
KEYWORDS antigen receptor; immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; variable region.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 345)
AUTHORS Domiatli-Saad,R. and Lipsky,P.E.
TITLE Staphylococcal enterotoxin A induces survival of VH3-expressing human B cells by binding to the VH region with low affinity
JOURNAL J. Immunol. 161 (3), 1257-1266 (1998)
MEDLINE 98349375
PUBMED 9686586
REFERENCE 2 (bases 1 to 345)
AUTHORS Domiatli-Saad,R.

TITLE Direct Submission
JOURNAL Submitted (23-JUN-1997) Domiatli-Saad R., Department of Internal Medicine, Harold C. Simmons Arthritis Research Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-8884, USA
FEATURES
source
1..345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/cell_type="B cell"
/tissue_type="peripheral blood"
/rearranged
1..345
V_region
/product="VH3, 3-15/DP-38"
BASE COUNT 75 a 83 c 116 g 70 t 1 others
ORIGIN
*Query Match 23.6%; Score 33; DB 9; Length 345;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTGT 43
|||||
Db 330 GCATGCTCCGGCGCCATGCGCGGATTGT 298

RESULT 13
LOCUS AR210351 163 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 52 from patent US 6387657.
ACCESSION AR210351
VERSION AR210351.1 GI:21512560
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 163)
AUTHORS Botstein,D.A., Cohen,R.L., Goddard,A.D., Gurney,A.L., Hillan,K.J., Lawrence,D.A., Levine,A.J., Pennica,D., Roy,M.Ann. and Wood,W.I.
TITLE WISP polypeptides and nucleic acids encoding same
JOURNAL Patent: US 6387657-A 52 14-MAY-2002;
FEATURES
source
1..163
/organism="unknown"
BASE COUNT 36 a 46 c 45 g 36 t
ORIGIN
Query Match 23.3%; Score 32.6; DB 6; Length 163;
Best Local Similarity 89.7%; Pred. No. 1;
Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTGTCA 49
|||||
Db 123 GCATGCTCCGGCGCCATGCGCGGATTACTAG 85

RESULT 14
LOCUS AC107595 155394 bp DNA linear HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-273u5, *** SEQUENCING IN PROGRESS ***
ACCESSION AC107595
VERSION AC107595.3 GI:21737300
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 155394)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,

Barbarta,J., Benton,J., Blmage,K., Blankenburg,R., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carlton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chave,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisti,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,K., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,F.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okunolu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Qutles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Soderren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Syatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vazquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,C.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

Direct Submission
 2 (bases 1 to 155394)
 Worley,K.C.
 Direct Submission
 Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 155394)
 Worley,K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced g1:16701904.

----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GOMP
 Center clone name: CH230-27315
 ----- Summary Statistics -----
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 103821 bases at least Q40
 Consensus quality: 108962 bases at least Q30
 Consensus quality: 113213 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a working draft sequence. It currently
 * consists of 66 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1631: contig of 1631 bp in length
2	1632
3	1731: gap of unknown length
4	1732
5	3156: contig of 1425 bp in length
6	3157
7	3256: gap of unknown length
8	3257
9	4636: contig of 1380 bp in length
10	4637
11	4736: gap of unknown length
12	4737
13	5349: contig of 1213 bp in length
14	5950
15	6049: gap of unknown length
16	6050
17	7220: contig of 1171 bp in length
18	7221
19	7320: gap of unknown length
20	7321
21	8496: contig of 1176 bp in length
22	8497
23	8596: gap of unknown length
24	8597
25	9730: contig of 1134 bp in length
26	9731
27	9830: gap of unknown length
28	9831
29	11519: contig of 1689 bp in length
30	11520
31	11619: gap of unknown length
32	11620
33	13277: contig of 1658 bp in length
34	13278
35	13377: gap of unknown length
36	13378
37	14984: contig of 1607 bp in length
38	14985
39	15084: gap of unknown length
40	15085
41	16292: contig of 1208 bp in length
42	16293
43	16392: gap of unknown length
44	16393
45	17616: contig of 1224 bp in length
46	17617
47	17716: gap of unknown length
48	17717
49	19494: contig of 1778 bp in length
50	19495
51	19594: gap of unknown length
52	19595
53	21600: contig of 2006 bp in length
54	21601
55	21700: gap of unknown length
56	21701
57	23028: contig of 1328 bp in length
58	23029
59	23129
60	24534: contig of 1406 bp in length
61	24535
62	24634: gap of unknown length
63	24635
64	26680: contig of 2046 bp in length
65	26681
66	26780: gap of unknown length
67	26781
68	29046: contig of 2266 bp in length
69	29047
70	29146: gap of unknown length
71	29147
72	30966: contig of 1820 bp in length
73	30967
74	31066: gap of unknown length
75	32514: contig of 1448 bp in length
76	32515
77	32614: gap of unknown length
78	32615
79	34738: contig of 2124 bp in length
80	34739
81	34838: gap of unknown length
82	34839
83	36825: contig of 1987 bp in length
84	36826
85	36925: gap of unknown length
86	38526
87	38569: contig of 1644 bp in length
88	38570
89	38669: gap of unknown length
90	40114: contig of 1445 bp in length
91	40115
92	40214: gap of unknown length
93	40215
94	42188: contig of 1974 bp in length
95	42189
96	42288: gap of unknown length
97	42189
98	43549: contig of 1261 bp in length
99	43550
100	43649: gap of unknown length
101	43650
102	45433: contig of 1784 bp in length
103	45434
104	45533: gap of unknown length
105	45534
106	47186: contig of 1653 bp in length
107	47187
108	47286: gap of unknown length
109	47287
110	48325: contig of 1039 bp in length
111	48326
112	48425: gap of unknown length
113	50047: contig of 1622 bp in length
114	50048
115	50147: gap of unknown length
116	50148
117	52155: contig of 2008 bp in length
118	52156
119	52255: gap of unknown length
120	52256
121	54715: contig of 2460 bp in length
122	54716
123	54815: gap of unknown length
124	54816
125	56189: contig of 1373 bp in length
126	56189
127	56288: gap of unknown length
128	56289
129	58334: contig of 2046 bp in length
130	58334

RESULT 15					
HV1371VRC					
LOCUS	HV1371VRC	278 bp	DNA	linear	PLN 07-JUL-2002
DEFINITION	H.vulgare BARE-1 long terminal repeat DNA; ID E1371VRC.				
ACCESSION	284567				
VERSION	284567.1	GI:2598520			
KEYWORDS	BARE-1 retrotransposon;	long terminal repeat.			
SOURCE	<i>Hordeum vulgare</i> .				
ORGANISM	<i>Hordeum vulgare</i> .				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
	Pooidae; Triticeae; Hordeum.				
	1 (bases 1 to 278)				
REFERENCE	Suontemi,A., Schmidt,D. and Schulman,A.H.				
AUTHORS	BARE-1 insertion site preferences and evolutionary conservation of				
TITLE					

JOURNAL	RNA and cDNA processing sites
MEDLINE	Journal 100 (1-3), 219-230 (1997)
PUBMED	98103449
REFERENCE	9440275
AUTHORS	2 (bases 1 to 278)
TITLE	Schulman,A.H.
JOURNAL	Direct Submission Submitted (24-JAN-1997) Schulman A.H., University of Helsinki, Institute of Biotechnology, P.O. Box 56, Biocentre 1, Viikinkaari 9, FIN-00014 Helsinki, Finland
FEATURES	location/Qualifiers
Source	1..278
	/organism="Hordeum vulgare"
	/db_xref="taxon:4513"
misc_feature	<1..89
LTR	/note="Flanking genomic sequence matches BARE-1a retrotransposon (Z17327) nt 1990-2078"
repeat_region	90..>222
	/note="Matches nt 1-89 of BARE-1 5' LTR, 5' terminus"
BASE COUNT	90 . 222
ORIGIN	/rnsposon="BARE-1"
	78 a 65 c 57 g 78 t
Query Match	23.1%: Score 32.4; DB 8; Length 278;
Best Local Similarity	97.1%: Pred. No. 1.3;
Matches	33: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
OY	11 GCATGCTCCGGCGGCATGCCGGGGCATTC 44
Dsb	 13 GCATGCTCCGGCGGCATGCCGGGGCATTC 46

Search completed: May 25, 2003, 19:41:04
Job time : 521.477 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:46:55 ; Search time 70.3404 Seconds
(without alignments)
4482.203 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
Sequence: 1 atagatacagcatgctccc.....gagctcgaatcgccctata 140

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140	100.0	140	21	AA245985
2	45	32.1	51	21	AA245973
3	39.6	28.3	477	21	AA245987
4	35.6	25.4	1949	22	AAK51591
5	35.4	25.3	1630	22	AA157798
6	35.2	25.1	1222	20	AAV69558
7	34.8	24.9	2597	22	AA158172
8	34.4	24.6	3186	22	AA157799
9	34.2	24.4	163	20	AAK76514

10	34.2	24.4	163	20	AAK55630	5' Cloning site re
11	34	24.3	2064	24	ABN84401	Human Kv4.3 potass
12	34	24.3	2121	24	ABN84400	Human Kv4.3 potass
13	33.8	24.1	484	18	AAK83364	Breast cancer tumo
14	33.8	24.1	484	19	AAV68904	DNA molecule encod
15	33.8	24.1	484	21	AAK80887	Human breast tumo
16	33.8	24.1	484	21	AAK46777	Human breast tumo
17	33.8	24.1	484	24	AAK59733	Breast tumour-spec
18	33.4	23.9	5506	24	ABK67846	Lettuce pest resis
19	33.4	23.9	13339	19	AAV44234	Soy bean SCN/SCS r
20	33.2	23.7	711	22	ABA06302	Infelious Bursel
21	33	23.6	3171	24	ABA93132	WSP 3' cloning si
22	32.6	23.3	163	20	AAK76515	3' cloning site re
23	32.6	23.3	163	20	AAK55631	Human ADNF-III PCR
24	32.6	23.3	837	19	AAV49815	Sense PCR product
25	32.6	23.3	837	21	AAV49815	Human ADNF-III PCR
26	32.6	23.3	850	19	AAV49816	Antisense PCR prod
27	32.6	23.3	850	21	AAV49816	Pear pectin methyl
28	32.2	23.0	700	24	AAK02962	Physioleca america
29	31.4	22.4	593	21	AAK97483	Soy bean SCN/SCS r
30	31.4	22.4	713	22	ABA06304	S. vititatum silk p
31	31.4	22.4	831	24	ABN86270	Human breast tumo
32	31	22.1	215	22	AAH57722	DNA encoding the h
33	31	22.1	420	19	AAV09802	DNA encoding novel
34	31	22.1	420	19	AAV09802	CDNA sequence of a
35	31	22.1	482	23	AAK71699	Breast cancer tumo
36	31	22.1	530	21	AAK47071	DNA molecule encod
37	31	22.1	597	18	AAK83374	Human breast tumo
38	31	22.1	597	19	AAV68913	Breast tumour-spec
39	31	22.1	597	21	AAK80897	Pythium venterpool
40	31	22.1	597	24	ABK46787	Human histiocyte-s
41	31	22.1	597	24	AAK59743	ECM3 gene promoter
42	31	22.1	919	21	AAK57434	Human poly nucleoti
43	31	22.1	931	17	AAK14818	
44	31	22.1	1205	18	AAK94303	
45	31	22.1	1265	22	AAK159000	

ALIGNMENTS

RESULT 1	
AA245985	
ID	AA245985 standard; cDNA; 140 Bp.
XX	
AC	AA245985;
XX	
DT	25-APR-2000 (first entry)
XX	
DE	Fragment of pJANG-deltaCys7Cys236 plasmid.
XX	
KW	Glutenin; seed storage protein; gluten; bread; pasta; noodle;
KW	breakfast cereal; snack food; cake; pastry; flour based sauce; grain;
KW	film; coating; adhesive; building material; packaging material; ss.
OS	Synthetic.
XX	
OS	Hordeum vulgare.
XX	
FT	key
FT	CDS
FT	Location/Qualifiers
FT	45..107
FT	/tag= a
PN	WO200002914-A1.
XX	
PD	20-JAN-2000.
XX	
PF	12-JUL-1999; 99MO-AU00563.
XX	
PR	10-JUL-1998; 98AU-0004604.
XX	
PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.
PA	(GOOD-) GOODMAN FIELDER LTD.
PA	(LIMA-) GRP LIMAGRAIN PACIFIC PRY LTD.

XX Appels R, Morell M, Bekes F, Tamas L;
 XX WPI: 2000-147597/13.
 DR P-PSDB; AAY54566.
 XX
 PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
 PT films, coatings, packing materials, adhesives and building materials -
 XX
 PS Disclosure; Fig 1: 76pp; English.
 XX
 CC The present sequence represents a fragment of the pJANG-deltaCys7Oys236
 CC plasmid. This plasmid encodes a barley glutenin, and is used in the
 CC course of the invention to produce modified glutenin. The specification
 CC describes a method for producing a modified glutenin or seed storage
 CC protein, by adding to the protein a domain that confers the ability to
 CC incorporate into gluten, or to bind a ligand or other macromolecule. The
 CC domain can be any domain that will bind ligands that may be useful in
 CC food preparation or in food composition, e.g. a domain that binds lipids
 CC or starch. The method is used for producing modified glutenins. Glutenins
 CC and seed-storage protein are useful for preparing food products such as
 CC leavened or unleavened breads, pasta, noodles, breakfast cereals, snack
 CC foods, cakes, pastries and foods containing flour based sauces. Glutenins
 CC and seed-storage protein are also useful for preparing nonfood products
 CC such as films, coatings, adhesives, building materials and packaging
 CC materials. Grain or parts of grain containing the modified glutenin and
 CC seed-storage protein is useful for preparing food products. The modified
 CC glutenins and seed-storage proteins are also useful as modifiers of
 CC food products in food industry.
 CC
 XX Sequence 140 BP: 34 A; 41 C; 36 G; 29 T; 0 other;

Query Match 100.0%; Score 140; DB 21; Length 140;
 Best Local Similarity 100.0%; Pred. No. 1,1e-38;
 Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ATGATATACAGCATGCTCCGCGCCGCCATGCGCGGATTTGTCATGACCACTAAACC 60
 DB 1 ATGATATACAGCATGCTCCGCGCCGCCATGCGCGGATTTGTCATGACCACTAAACC 60
 OY 61 CTTCCAGCGTCCCGCCACAGCTTCATGATGAGTATGCTAGGATCCGGGTACC 120
 DB 61 CTTCCAGCGTCCCGCCACAGCTTCATGATGAGTATGCTAGGATCCGGGTACC 120
 OY 121 GAGCTCGAATTCGCCCTATA 140
 DB 121 GAGCTCGAATTCGCCCTATA 140

RESULT 2
 AAZ45973/C
 ID AAZ45973 standard; DNA: 51 BP.

XX AAZ45973;
 AC
 XX 25-APR-2000 (first entry)
 DT
 XX PCR primer used to amplify a 447 bp fragment of C hordein gene.
 DE
 XX C hordein gene; glutenin; seed storage protein; gluten; bread; pasta;
 KW noodle; breakfast cereal; snack food; cake; pastry; flour based sauce;
 KW film; coating; adhesive; building material; packaging material; grain;
 KW PCR primer; ss.
 XX
 OS Hordeum vulgare.
 XX
 PN WO200002914-A1.
 XX
 PD 20-JAN-2000.
 XX
 PF 12-JUL-1999; 99WO-AU00563.
 XX
 PR 10-JUL-1998; 98AU-0004604.

XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (GOOD-) GOODMAN FIELDER LTD.
 PA (LIMA-) GRP LIMABRAIN PACIFIC PTY LTD.
 XX
 PT Appels R, Morell M, Bekes F, Tamas L;
 DR WPI: 2000-147597/13.
 XX
 PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
 PT films, coatings, packing materials, adhesives and building materials -
 XX
 PS Disclosure; Page 11: 76pp; English.
 XX
 CC PCR primers AAZ45972-73 were used to amplify a 477 bp fragment
 CC (approximately 2/3 of the whole gene) of the barley C hordein gene. This
 CC gene encodes a glutenin, and was modified according to the method of
 CC the invention. The specification describes a method for producing a
 CC modified glutenin or seed storage protein, by adding to the protein a
 CC domain that confers the ability to incorporate into gluten, or to bind
 CC a ligand or other macromolecule. The domain can be any domain that will
 CC bind ligands that may be useful in food preparation or in food
 CC composition, e.g. a domain that binds lipids or starch. The method is
 CC used for producing modified glutenins. Glutenins and seed-storage
 CC protein are useful for preparing food products such as leavened or
 CC unleavened breads, pasta, noodles, breakfast cereals, snack foods, cakes,
 CC pastries and foods containing flour based sauces. Glutenins and
 CC seed-storage protein are also useful for preparing nonfood products such
 CC as films, coatings, adhesives, building materials and packaging
 CC materials. Grain or parts of grain containing the modified glutenin and
 CC seed-storage protein is useful for preparing food products. The modified
 CC glutenins and seed-storage proteins are also useful as modifiers of food
 CC products in food industry.
 CC
 XX Sequence 51 BP: 11 A; 11 C; 15 G; 14 T; 0 other;

Query Match 32.1%; Score 45; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 5.6e-06;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 69 GTCCGCCACAGCTTCATGATGAGTATGCTAGGATCC 113
 DB 45 GTCCGCCACAGCTTCATGATGAGTATGCTAGGATCC 1

RESULT 3
 AAZ45987
 ID AAZ45987 standard; CDNA: 477 BP.

XX AAZ45987;
 AC
 XX 25-APR-2000 (first entry)
 DT
 XX cDNA encoding a synthetically truncated C hordein protein (glutenin).
 DE
 XX C hordein gene; glutenin; seed storage protein; gluten; bread; pasta;
 KW noodle; breakfast cereal; snack food; cake; pastry; flour based sauce;
 KW film; coating; adhesive; building material; packaging material; grain;
 KW ss.
 XX
 OS Hordeum vulgare.
 XX
 OS Synthetic.
 XX
 FH Key
 FT CDS
 FT 1..477
 FT Location/Qualifiers
 FT /tag= a
 FT /transl_except= (pos: 58..60, aa: Gln)
 FT /transl_except= (pos: 313..315, aa: Gln)
 FT /transl_except= (pos: 316..318, aa: Gln)
 FT /transl_except= (pos: 319..321, aa: Pro)
 FT /transl_except= (pos: 322..324, aa: Phe)
 FT /transl_except= (pos: 337..339, aa: Gln)
 FT /transl_except= (pos: 340..342, aa: Gln)

```

FT      /transl_except= (pos: 343..345, aa: Pro)
FT      /transl_except= (pos: 346..348, aa: Phe)
FT      /transl_except= (pos: 349..351, aa: Pro)
FT      /transl_except= (pos: 352..354, aa: Gln)
FT      /transl_except= (pos: 355..357, aa: Gln)
FT      /transl_except= (pos: 358..360, aa: Pro)
XX
XX      WO200002914-A1.
XX
XX      20-JAN-2000.
XX
XX      12-JUL-1999: 99WO-AU00563.
XX
XX      10-JUL-1998: 98AU-0004604.
XX
XX      (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX      (GOOD-) GOODMAN FIELDER LTD.
XX      (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
XX
XX      Appels R, Morell M, Bekes F, Tamas L;
XX      WPI: 2000-147597/13.
XX      P-PSDB: AAY54568.
XX
XX      Modifying glutenin or seed-storage protein for preparing foodstuffs,
XX      films, coatings, packing materials, adhesives and building materials -
XX
XX      Disclosure: Fig 3; 76pp; English.
XX
XX      The present sequence encodes a synthetically truncated barley C
XX      hordein protein, which is a glutenin. The protein is designated
XX      CC AMG-deltaCys7Cys236, and is modified, according to the method of
XX      CC the invention. The specification describes a method for producing a
XX      CC modified glutenin or seed storage protein, by adding to the protein a
XX      CC domain that confers the ability to incorporate into gluten, or to bind
XX      CC a ligand or other macromolecule. The domain can be any domain that will
XX      CC bind ligands that may be useful in food preparation or in food
XX      CC composition, e.g. a domain that binds lipids or starch. The method is
XX      CC used for producing modified glutenins. Glutenins and seed-storage protein
XX      CC are useful for preparing food products such as leavened or unleavened
XX      CC breads, pasta, noodles, breakfast cereals, snack foods, cakes, pastries
XX      CC and foods containing flour based sauces. Glutenins and seed-storage
XX      CC protein are also useful for preparing nonfood products such as films,
XX      CC coatings, adhesives, building materials and packaging materials. Grain
XX      CC or parts of grain containing the modified glutenin and seed-storage
XX      CC protein is useful for preparing food products. The modified glutenins and
XX      CC seed-storage proteins are also useful as modifiers of food products in
XX      CC food industry.
XX
XX      Sequence 477 BP: 162 A; 186 C; 43 G; 86 T; 0 other:
XX
XX      Query Match      28.3%: Score 39.6; DB 21: Length 477;
XX      Best Local Similarity 83.3%: Pred. No. 0.00076;
XX      Matches 45; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
XX
OY      54 CTTAAACCCCTGCAGCGCCGCCAACAAGCTCATGATGATGAGATGCTTAG 107
OY      |||||
OY      424 CACACACCTCAACCAAGTCCGCCACACAGCTTCATGATGAGATGCTTAG 477
XX
XX      RESULT 4
XX      AAK51591
XX      ID AAK51591 standard; cDNA: 1949 BP.
XX
XX      AAK51591:
XX
XX      06-NOV-2001 (first entry)
XX
XX      Human polynucleotide SEQ ID NO 136.
XX
XX      Human: cytokine; cell proliferation; cell differentiation; gene therapy;
XX      KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX      tissue growth factor; immunomodulatory; cancer; leukaemia;

```

```

KW      nervous system disorder; arthritis; inflammation; ss.
XX
XX      Homo sapiens.
OS
XX      WO200157190-A2.
XX
XX      09-AUG-2001.
XX
XX      05-FEB-2001: 2001WO-US04098.
XX
XX      03-FEB-2000: 2000US-0496914.
XX      27-APR-2000: 2000US-0560875.
XX      20-JUN-2000: 2000US-0598075.
XX      19-JUL-2000: 2000US-0620325.
XX      01-SEP-2000: 2000US-0634936.
XX      15-SEP-2000: 2000US-0663561.
XX      20-OCT-2000: 2000US-0693325.
XX      30-NOV-2000: 2000US-0728422.
XX
XX      (HYSE-) HYSEQ INC.
XX
XX      Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX      PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX      PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX      WPI: 2001-476283/51.
XX      P-PSDB: AAM78458.
XX
XX      Nucleic acids encoding polypeptides with cytokine-like activities,
XX      PT useful in diagnosis and gene therapy -
XX
XX      Claim 1: Page 816-818; 6221pp; English.
XX
XX      The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX      CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX      CC cytokine, cell proliferation or cell differentiation or which may induce
XX      CC production of other cytokines in other cell populations. The
XX      CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX      CC peptide therapy. The polypeptides have various cytokine-like activities,
XX      CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX      CC activity, tissue growth factor activity, immunomodulatory activity and
XX      CC activin/inhibin activity and may be useful in the diagnosis and/or
XX      CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX      CC inflammation.
XX      CC Note: Records for SEQ ID NO 2110 (AAK52561), 2111 (AAK52582) and 3666
XX      CC (AAM80020) are omitted as the relevant pages from the sequence listing
XX      CC were missing at the time of publication.
XX
XX      Sequence 1949 BP: 484 A; 434 C; 517 G; 514 T; 0 other:
XX
XX      Query Match      25.4%: Score 35.6; DB 22: Length 1949;
XX      Best Local Similarity 75.9%: Pred. No. 0.027;
XX      Matches 44; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
XX
OY      11 GCATGCTCCCGCGCGCATGCGCGGATTCATGACGCAACTTAACCTTGCAGC 68
OY      |||||
OY      10 GCATGCTCCCGCGCGCATGCGCGGATTTTCCGCGGTCGACGATTTTCGACC 67
XX
XX      RESULT 5
XX      AA157798
XX      ID AA157798 standard; cDNA: 1630 BP.
XX
XX      AA157798:
XX
XX      22-OCT-2001 (first entry)
XX
XX      Human polynucleotide SEQ ID NO 1.
XX
XX      Human: neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX      KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX      Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX      amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

```

KM		chemokine; chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW		leukemia; ss.
XX		
OS	Homo sapiens.	
PN	WO200153312-AL.	
PD		
XX	26-JUL-2001.	
XX		
PF	26-DEC-2000; 2000MO-US34263.	
XX		
PR	21-JAN-2000; 2000US-0488725.	
PR	25-APR-2000; 2000US-0552317.	
PR	09-JUL-2000; 2000US-0598042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
PR	29-NOV-2000; 2000US-0727344.	
XX		
PA	(HXB-) HXSEQ INC.	
PI	Tang Y.T., Liu C., Asundi V., Chen R., Ma Y., Qian X.B., Ren F., Wang D.;	
P1	Wang Z., Wehrman T., Xu C., Xue A.J., Yang Y., Zhang J.;	
PI	Zhao Q.A., Zhou P., Goodrich R., Dimanac RT;	
XX		
DR	WPI: 2001-442253/47.	
PS	P-PSDB: AAM38642.	
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
XX	such as central nervous system injuries -	
XX	Claim 1; SEQ ID NO 1; 10078pp; English.	
XX		
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and	
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,	
CC	immunosuppressant and cyostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukemias and	
CC	C.N.S. disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
SQ	Sequence 1630 BP; 392 A; 422 C; 438 G; 378 T; 0 other:	
QY	Query Match 25.3%; Score 35.4; DB 22; Length 1630;	
QY	Best Local Similarity 79.2%; Pred. No. 0.03;	
QY	Matches 42; Conservative 0; Mismatches 11; Indels 0; Gaps 0	
DB	11 GCATCCTCCGCGCCCATGGCCGGAGATTGCATGACGAACAACCCCT 63 10 GCATCCTCCGCGCCCATGGCCGGAGATTTCGCCGATTCGATT 62	
RESULT 6		
ID	AAV69558/c	
XX	AAV69558 standard; DNA; 1222 BP.	
AC	AAV69558;	
DT	15-MAR-1999 (first entry)	
DE	Soil derived peptide synthase clone ps3 DNA.	
XX	peptide synthase; soil; lichen; antibiotic biosynthesis; humus;	

KM		therapeutic; immunosuppressor; antitumour agent; pathogen;
KM		genetic diversity; ss.
XX	OS	Unknown.
XX	PN	W09853097-A2.
XX	PD	26-NOV-1998.
XX	PF	21-MAY-1998; 98WO-CA00488.
XX	PR	22-MAY-1997; 97US-0861774.
XX	PA	(TERR-) TERRAGEN DIVERSITY INC.
XX	PI	Miao YPM, Seow KT, Waters B, Yap WH;
XX	DR	WPI; 1999-070158/06.
XX	P	P-PSDB; AAM82710.
PT		New degenerate primers - used for recovering antibiotic biosynthetic DNA from soil/lichen material
PS		Claim 17; Page 80; 98pp; English.
CC		This sequence encodes a peptide synthase clone, isolated from soil.
CC		This protein is used in a method for the recovery of antibiotic biosynthetic DNA from humic materials or lichen. The PCR products of the invention have the potential to be used as therapeutic molecules including antibiotics, immunosuppressors and antitumour agents. The method allows access to the reservoir of genetic diversity in soil pathogens micro-organisms in order to find new antibiotics. It also allows access to novel biosynthetic genes/enzymes that can be used to produce antibiotics or produce specific compounds, enzymatically, in vitro.
SQ		Sequence 1222 BP; 272 A; 361 C; 319 G; 270 T; 0 other:
Query Match	25.1%;	Score 35.2; DB 20; Length 1222;
Best Local Similarity	92.5%;	Pred No. 0.033; 3; Indels 0; Gaps 0;
Matches 37; Conservative	0;	Mismatches 3;
OY	11 GCATGCTCCCGGCCGCATGGCGCGGATTGTGATGAGG 50	
DB	1176 GCATGCTCCCGGCCGCATGGCGCGGATTGTGACAGCAGC 1137	
RESULT 7		
AAI58172/c		
ID	AAI58172 standard; cDNA; 2597 BP.	
XX	AC	AAI58172;
DT	22-OCT-2001 (first entry)	
DE	Human polynucleotide SEQ ID NO 375.	
XX		
XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KM	peripheral nervous system; neuropathy; central nervous system; CNS;	
KM	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KM	amyotrophic lateral sclerosis; Shy-Draeger Syndrome; chemotactic;	
KM	chemokinesis; thrombolytic; drug screening; arthritis; inflammation;	
XX	leukaemia; ss.	
XX		
OS	Homo sapiens.	
PN	W0200153312-A1.	
XX	26-JUL-2001.	
PD		
PF	26-DEC-2000; 2000WO-US34263.	
XX		
PR	21-JAN-2000; 2000US-0488725.	

PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HYSEQ-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB: AAM39016.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1: SEQ ID NO 375; 10078bp: English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 2597 BP; 733 A; 607 C; 690 G; 567 T; 0 other;
 XX
 XX Query Match 24.9%; Score 34.8; DB 22; Length 2597;
 XX Best Local Similarity 84.8%; Pred. No. 0.055; 7; Indels 0; Gaps 0;
 XX Matches 39; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 11 GCATGCTCCCGCGCCGATGCGGATGTCATGAGGCACTA 56
 DB 2561 GCATGCTCCCGCGCCGATGCGGATGTCATGAGGCACTA 2516
 RESULT 8
 ID AA157799 standard; CDNA: 3186 BP.
 AC AA157799;
 XX
 XX 22-OCT-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 2.
 XX
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX

PF 26-DEC-2000; 2000MO-US34263.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HYSEQ-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB: AAM38643.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1: SEQ ID NO 2; 10078bp: English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 3186 BP; 695 A; 1058 C; 894 G; 539 T; 0 other;
 XX
 XX Query Match 24.6%; Score 34.4; DB 22; Length 3186;
 XX Best Local Similarity 86.4%; Pred. No. 0.08;
 XX Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 11 GCATGCTCCCGCGCCGATGCGGATGTCATGAGGCAAC 54
 DB 20 GCATGCTCCCGCGCCGATGCGGATGTCATGAGGCAAC 63
 RESULT 9
 ID AAX76514 standard; DNA: 163 BP.
 AC AAX76514;
 XX
 XX 06-AUG-1999 (first entry)
 DE WISP 5' cloning site region SEQ ID NO:51.
 XX
 XX WNT-1 induced secreted protein; WISP-1; WISP-2; WISP-3; CTGF; tumour;
 KM connective tissue growth factor; cancer; melanoma; arteriosclerosis;
 KM leukaemia; lymphoid malignancy; haematopoiesis-related disorder;
 KM tissue-growth disorder; skin disorder; desmoplasia; fibrotic lesion;
 KM kidney disorder; bone-related disorder; osteoporosis; trauma; burn;
 KM connective tissue disorder; catabolic state; inflammation;
 KM testicular-related disorder; angiogenesis; immunological disorder; ss.
 XX
 OS Synthetic.
 XX

PN W09921998-A1.
 XX
 PD 06-MAY-1999.
 XX
 PF 29-OCT-1998; 98WO-US22991.
 XX
 PR 14-APR-1998; 98US-0081695.
 PR 29-OCT-1997; 97US-0063704.
 PR 03-FEB-1998; 98US-0073612.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Botstein DA, Cohen RL, Goddard A, Gurney AL, Hillan K;
 PI Lawrence DA, Levine AJ, Pennica D, Roy MA, Wood WI;
 DR WPI; 1999-337420/28.
 XX
 PT New isolated Wnt-1 induced secreted polypeptides, WISP-1, 2 and 3
 PS
 XX Example 1; Page 203; 284pp: English.
 PS
 XX The present invention describes Wnt-1 induced secreted polypeptides,
 CC WISP-1, 2 and 3. The novel WISP polypeptides, designated WISP-1, WISP-2
 CC and WISP-3 have homology to connective tissue growth factor (CTGF).
 CC Products from the present invention can be used to treat WISP-related
 CC disorders such as breast, ovarian, and colon cancer or melanoma. The
 CC products can be used to treat arteriosclerosis. The products can also be
 CC used to treat other diseases e.g. benign and malignant tumors,
 CC leukaemia and lymphoid malignancies, neuronal, glial, astrocytal,
 CC hypothalamic and other glandular, macrophagal, epithelial, stromal, and
 CC blastocoeal disorders, haematopoiesis-related disorders, tissue-growth
 CC disorders, skin disorders, desmoplasia, fibrotic lesions, kidney
 CC disorders, bone-related disorders such as osteoporosis, trauma such as
 CC burns, incisions, and other wounds, connective tissue disorders,
 CC catabolic states, testicular-related disorders, and inflammatory,
 CC angiogenic and immunologic disorders including arteriosclerosis. The
 CC products can also be used for detection and diagnosis especially of
 CC individuals with neoplastic cell growth or proliferation. The products
 CC can be used in the production of transgenic or knock-out animals.
 CC Antibodies can be used to induce death in WISP-1, 2 or 3 overexpressing
 CC cells.
 CC
 SQ Sequence 163 BP; 35 A; 45 C; 46 G; 37 T; 0 other;
 Query Match 24.4%; Score 34.2; DB 20; Length 163;
 Best Local Similarity 92.3%; Pred. No. 0.041; Mismatches 3; Indels 0; Gaps 0;
 Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 11 GCATGCTCCCGCGCCGATGCGCGGATGTGTCATGAG 49
 Db 41 GCATGCTCCCGCGCGCCGATGCGCGGATGTATCACTAG 79
 RESULT 10
 AAX55630
 ID AAX55630 standard; DNA; 163 BP.
 XX
 AC AAX55630;
 XX
 DT 26-JUL-1999 (first entry)
 XX
 DE 5' cloning site region of plasmid pGEM-T.
 XX
 KW Human; murine; clone 65; clone 320; Wnt; arteriosclerosis; malignancy;
 KW melanoma; cancer; breast; ovary; colon; tumor; cardiac; renal;
 KW inflammation; angiogenesis; immunological disorder; ss.
 XX
 OS Synthetic.
 XX
 PN W09921999-A2.
 PD 06-MAY-1999.

PF 29-OCT-1998; 98WO-US22992.
 XX
 PR 04-FEB-1998; 98US-0073612.
 PR 29-OCT-1997; 97US-0063704.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Levine AJ, Pennica D;
 DR WPI; 1999-326705/27.
 XX
 PT Human and murine genes induced by Wnt and derived polypeptides
 PS
 XX Examples; Fig 9; 114pp: English.
 PS
 XX The invention relates to human and murine cDNA clones 65 and 320,
 CC encoding polypeptides that are induced by Wnt. Clone 65 and 320
 CC polypeptides, and their antagonists, are used to treat disorders
 CC associated with these polypeptides, e.g. arteriosclerosis, malignancies
 CC (particularly melanoma or cancer of the breast, ovary and colon), but
 CC also benign tumors, cardiac, renal, catabolic or many other types of
 CC disease. Inflammation, angiogenesis and immunological disorders, more
 CC generally to induce death of Wnt-induced cells. Antibodies against the
 CC polypeptides are used as therapeutic antagonists and as diagnostic
 CC immunoassay or affinity-purification reagents. Clone 65 and 320 nucleic
 CC acids are used as hybridization probes or primers (to detect related
 CC sequences); for chromosome and gene mapping; to generate antisense
 CC transgenic or 'knockout' animals (for development of, and screening for,
 CC drugs). The polypeptides are used to raise (or purify) specific
 CC antibodies; as immunoassay reagents and in drug screens.
 CC
 SQ Sequence 163 BP; 35 A; 45 C; 46 G; 37 T; 0 other;
 Query Match 24.4%; Score 34.2; DB 20; Length 163;
 Best Local Similarity 92.3%; Pred. No. 0.041; Mismatches 3; Indels 0; Gaps 0;
 Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 11 GCATGCTCCCGCGCCGATGCGCGGATGTGTCATGAG 49
 Db 41 GCATGCTCCCGCGCGCCGATGCGCGGATGTATCACTAG 79
 RESULT 11
 ABN84401/c
 ID ABN84401 standard; cDNA; 2064 BP.
 XX
 AC ABN84401;
 XX
 DT 01-OCT-2002 (first entry)
 XX
 DE Human Kv4.3 potassium channel (short form) cDNA.
 XX
 KW Kv4.3; potassium channel; human; Alzheimer's disease;
 KW heart disease; neurotropic; neuroprotective; cardiac; gene therapy;
 KW gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH CDS 73..1983
 FT /*tag= a
 FT /product= "Kv4.3"
 XX
 XX US6395477-B1.
 XX
 PD 28-MAY-2002.
 XX
 PF 23-OCT-1998; 98US-0178109.
 XX
 PR 23-OCT-1998; 98US-0178109.
 XX
 PA (AMHP) AMERICAN HOME PROD CORP.

XX PI Cockett MI, DILKS DW, Ling HC, Sokol PT;
 XX DR WPI: 2002-556093/59.
 XX DR P-PSDB: ABB79585.

XX PT New isolated polynucleotide encoding human Kv4.3 potassium channel
 PT polypeptide, useful as probe in a diagnostic method for detecting
 PT nucleic acid encoding human Kv4.3, and for treating Alzheimer's and
 PT heart diseases

XX PS Claim 7; Column 21-22; 19pp; English.

XX CC The present sequence is that of cDNA encoding the short isoform
 CC of novel human potassium channel Kv4.3. 2 isoforms of human Kv4.3
 CC have been identified. One form is full-length (hKv4.3 long) while
 CC the second form has a deletion of 19 amino acids in the carboxy
 CC domain after the predicted sixth transmembrane domain (hKv4.3 short).
 CC Human heart primarily expresses hKv4.3 long, whereas human brain
 CC contains both forms. To obtain the present hKv4.3 short cDNA, PCR
 CC amplification was performed using primers that flanked the 57 bp
 CC insert in hKv4.3 long. The invention provides hKv4.3 polypeptides,
 CC polynucleotides, and methods for producing these polynucleotides.
 CC The Kv4.3 polypeptides and polynucleotides are useful in the
 CC diagnosis, treatment and screening of human diseases relating to an
 CC excess or deficiency of hKv4.3 activity, including Alzheimer's
 CC disease and heart disease.

XX S0 Sequence 2064 BP; 446 A; 678 C; 545 G; 395 T; 0 other;

Query Match 24.3%; Score 34; DB 24; Length 2064;
 Best Local Similarity 88.1%; Pred. No. 0.098;
 Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 11 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCA 52
 Db 2015 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCA 1974

RESULT 12
 ABB84400/c
 ID ABB84400 standard; cDNA: 2121 BP.
 AC ABB84400;
 XX 01-OCT-2002 (first entry)
 DE Human Kv4.3 potassium channel (long form) cDNA.
 XX KW Kv4.3; potassium channel; human; Alzheimer's disease;
 KW heart disease; nootropic; neuroprotective; cardiac; gene therapy;
 KW gene; ss.
 XX OS Homo sapiens.
 XX FT Key Location/Qualifiers
 FT CDS 73..2040
 FT /*tag= a
 FT /product= "Kv4.3"

US6395477-B1.
 28-MAY-2002.
 23-OCT-1998; 98US-0178109.
 23-OCT-1998; 98US-0178109.
 23-OCT-1998; 98US-0178109.
 (AMHP) AMERICAN HOME PROD CORP.
 Cockett MI, DILKS DW, Ling HC, Sokol PT;
 WPI: 2002-556093/59.

XX DR P-PSDB: ABB79584.
 XX PT New isolated polynucleotide encoding human Kv4.3 potassium channel
 PT polypeptide, useful as probe in a diagnostic method for detecting
 PT nucleic acid encoding human Kv4.3, and for treating Alzheimer's and
 PT heart diseases

XX PS Claim 5; Column 15-18; 19pp; English.

XX CC The present sequence is that of cDNA encoding the long isoform of
 CC novel human potassium channel Kv4.3. To obtain the cDNA,
 CC oligonucleotides based on the published rat sequence were used to
 CC screen a whole heart cDNA library. A 511 bp fragment was obtained
 CC and used as a probe to rescreen the library. The resulting clones
 CC lacked the extreme 5' and 3' coding regions, and 5' and 3' RACE
 CC was therefore used to amplify these sequences from a human
 CC brainstem cDNA library. 2 isoforms of human Kv4.3 were identified.
 CC One form is full-length (hKv4.3 long) while the second form has a
 CC deletion of 19 amino acids in the carboxy domain after the
 CC predicted sixth transmembrane domain (hKv4.3 short). Human heart
 CC primarily expresses hKv4.3 long, whereas human brain contains both
 CC forms. The invention provides Kv4.3 polypeptides, polynucleotides,
 CC and methods for producing these polynucleotides. The Kv4.3
 CC polypeptides and polynucleotides are useful in the diagnosis,
 CC treatment and screening of human diseases relating to an excess or
 CC deficiency of hKv4.3 activity, including Alzheimer's disease and
 CC heart disease.

XX S0 Sequence 2121 BP; 457 A; 694 C; 557 G; 413 T; 0 other;

Query Match 24.3%; Score 34; DB 24; Length 2121;
 Best Local Similarity 88.1%; Pred. No. 0.099;
 Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 11 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCA 52
 Db 2072 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCA 2031

RESULT 13
 AAX83364
 ID AAX83364 standard; cDNA: 484 BP.
 AC AAX83364;
 XX 31-AUG-1999 (first entry)
 DE Breast cancer tumour specific clone #118.
 XX KW Breast cancer; tumour; gene expression; genome; diagnosis; mammal;
 KW human endogenous retrovirus; vaccine; ss.
 XX OS Homo sapiens.
 XX FT W09725426-A2.
 FT 17-JUL-1997.
 XX 10-JAN-1997; 97WO-US00485.
 XX 20-AUG-1996; 96US-0700014.
 XX 11-JAN-1996; 96US-0585392.
 XX (CORI-) CORIXA CORP.
 XX FT Frudakis TN, Reed SG, Smith JM.
 XX WPI: 1997-372865/34.
 XX Breast cancer-related DNA from retrovirus antigen (s) - useful for
 XX diagnosis and treatment of breast cancer
 XX Claim 1; Page 150-151; 221pp; English.

XX Sequences AAX83201-X83285 and AAX83331-X83415 represent novel breast
 CC cancer tumour specific clones which are expressed from a genomic region
 CC containing a human endogenous retrovirus (AAX83330). Detection of
 CC the clone sequences allows determination of the presence of breast
 CC cancer in a mammal. Progression of breast cancer can be monitored by
 CC detecting the level of clone expression. Polypeptides encoded by the
 CC clones can be used in vaccines to inhibit or prevent breast cancer.

XX Sequence 484 BP; 101 A; 138 C; 126 G; 118 T; 1 other;

Query Match 24.1%; Score 33.8; DB 18; Length 484;

Best Local Similarity 75.9%; Pred. No. 0.077; Mismatches 13; Indels 0; Gaps 0;

DB 11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAACCCCTTG 64
 25 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAACCCCTTG 78

RESULT 14

AAV68904 standard; DNA: 484 BP.

AAV68904;

22-JAN-1999 (first entry)

DNA molecule encoding a breast tumour specific polypeptide #96.

Human; breast cancer; breast tumour tissue; diagnosis; treatment;
 vaccine; epitope; endogenous; retroviral element; ss.

Homo sapiens.

WO945328-A2.

15-OCT-1998.

09-APR-1998; 98WO-0506939.

11-DEC-1997; 97US-0991789.

09-APR-1997; 97US-0838762.

(CORI-) CORIXA CORP.

Frudakis TN, Reed SG, Smith JM;

WPI: 1998-557473/47.

New DNA sequences isolated from endogenous human retroviral element
 and related vectors, transformed cells, proteins and antibodies,
 useful for diagnosis, treatment and prevention of breast cancer.

Claim 1; Page 93; 173pp; English.

AAV68800 to AAV68998 represent nucleotide sequences which encode human
 breast tumour specific polypeptides. Detection or measurement of
 human breast tumour specific polypeptides and nucleotide sequences,
 or the corresponding RNA in a sample, is used for diagnosis and
 monitoring of breast cancer. Human breast tumour specific polypeptides
 and nucleotide sequences, and the vectors containing the DNAs, are also
 useful in vaccines for inhibiting development (for prevention or
 therapy) of breast cancer. The polypeptides may also be used to
 raise monoclonal antibodies, used as immunoassay reagents.

Sequence 484 BP; 101 A; 138 C; 126 G; 118 T; 1 other;

Query Match 24.1%; Score 33.8; DB 19; Length 484;

Best Local Similarity 75.9%; Pred. No. 0.077; Mismatches 13; Indels 0; Gaps 0;

11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAACCCCTTG 64

DB 11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAACCCCTTG 78

RESULT 15

AAC80887 standard; cDNA: 484 BP.

AAC80887;

13-FEB-2001 (first entry)

Human breast tumour-specific cDNA SEQ ID NO: 175.

Human; breast tumour-specific antigen; cytostatic; vaccine;
 breast cancer; B18Ag1; B11Ag1; B15Ag1; ss.

Homo sapiens.

WO200061753-A2.

19-OCT-2000.

07-APR-2000; 2000WO-0509312.

09-APR-1999; 99US-0289198.

28-OCT-1999; 99US-0429755.

23-MAR-2000; 2000US-0534825.

(CORI-) CORIXA CORP.

Frudakis TN, Smith JM, Reed SG, Misher LE, Retter MW, Dillon DC;

WPI: 2000-628403/60.

An isolated polypeptide comprising an immunogenic portion of a breast
 tumor protein used for inhibiting the development of cancer, especially
 breast cancer, and monitoring cancer progression in a patient.

Claim 4; Page 135; 187pp; English.

The present sequence is given in a specification relating to compositions
 and methods for the treatment and diagnosis of breast cancer. Nucleotide
 sequences that are preferentially expressed in breast tumour tissue, and
 the polypeptides encoded by such nucleotide sequences, are used in
 compositions and vaccines to inhibit the development of cancer,
 especially breast cancer. The progression of a cancer may be monitored by
 carrying out detection of tumour-specific antigens at subsequent time
 points and comparing the results from the different time points.
 CD4 and/or CD8+ T-cells isolated from the cancer patient may be treated
 with tumour-specific polypeptides, polynucleotides encoding the
 polypeptides or antigen presenting cells expressing the polypeptides. The
 cells are then administered to the patient to inhibit development of
 cancer.

Sequence 484 BP; 101 A; 138 C; 126 G; 118 T; 1 other;

Query Match 24.1%; Score 33.8; DB 21; Length 484;

Best Local Similarity 75.9%; Pred. No. 0.077; Mismatches 13; Indels 0; Gaps 0;

DB 11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAACCCCTTG 64
 25 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAACCCCTTG 78

Search completed: May 25, 2003, 19:02:42
 Job time : 71.3404 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:55:25 ; Search time 15.2026 Seconds

(without alignments)
2824.173 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
1 atagatacagatgcctccc.....gagctcgaattgcgcctata 140

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PTDUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.2	25.1	1222	4 US-08-861-774E-83	Sequence 83, Appl
2	34.2	24.4	163	4 US-09-182-145-51	Sequence 51, Appl
3	34	24.3	2064	4 US-09-178-109-3	Sequence 1, Appl
4	34	24.3	2121	4 US-09-178-109-1	Sequence 1, Appl
5	33.8	24.1	484	4 US-08-991-789A-175	Sequence 175, App
6	33.8	24.1	484	4 US-09-062-451-175	Sequence 175, App
7	33.8	24.1	484	4 US-09-598-326-175	Sequence 175, App
8	33.4	23.9	5506	4 US-09-004-838-93	Sequence 93, Appl
9	32.6	23.3	163	4 US-09-182-145-52	Sequence 92, Appl
10	31	22.1	419	4 US-09-214-095D-97	Sequence 97, Appl
11	31	22.1	420	4 US-09-214-095D-111	Sequence 111, App
12	31	22.1	597	4 US-08-991-789A-185	Sequence 185, App
13	31	22.1	597	4 US-09-062-451-185	Sequence 185, App
14	31	22.1	597	4 US-09-598-326-185	Sequence 185, App
15	30.8	22.0	500	4 US-08-991-789A-167	Sequence 167, App
16	30.8	22.0	500	4 US-09-062-451-167	Sequence 167, App
17	30.8	22.0	500	4 US-09-598-326-167	Sequence 167, App
18	30.4	21.7	488	4 US-08-991-789A-180	Sequence 180, App
19	30.4	21.7	488	4 US-09-062-451-180	Sequence 180, App
20	30.4	21.7	488	4 US-09-598-326-180	Sequence 180, App
21	30.4	21.7	507	4 US-08-991-789A-182	Sequence 182, App
22	30.4	21.7	507	4 US-09-062-451-182	Sequence 182, App
23	30.4	21.7	507	4 US-09-598-326-182	Sequence 182, App
24	30	21.4	360	4 US-09-214-095D-91	Sequence 91, Appl
25	30	21.4	5505	4 US-09-004-838-93	Sequence 93, Appl
26	29.8	21.3	5829	4 US-09-605-785-473	Sequence 473, App
27	29.8	21.3	5829	4 US-09-439-513-473	Sequence 473, App

28	29.4	21.0	597	4 US-08-991-789A-186	Sequence 186, App
29	29.4	21.0	597	4 US-09-062-451-186	Sequence 186, App
30	29.4	21.0	597	4 US-09-598-326-186	Sequence 186, App
31	29	20.7	93	4 US-08-986-659B-15	Sequence 15, Appl
32	29	20.7	165	2 US-08-484-956-49	Sequence 49, Appl
33	29	20.7	165	2 US-08-757-653-49	Sequence 49, Appl
34	29	20.7	165	2 US-08-520-946-49	Sequence 49, Appl
35	29	20.7	206	1 US-08-254-359A-32	Sequence 32, Appl
36	29	20.7	206	1 US-08-471-066B-32	Sequence 32, Appl
37	29	20.7	206	2 US-08-484-956-32	Sequence 32, Appl
38	29	20.7	206	2 US-08-484-956-50	Sequence 50, Appl
39	29	20.7	206	2 US-08-757-653-32	Sequence 32, Appl
40	29	20.7	206	2 US-08-757-653-50	Sequence 50, Appl
41	29	20.7	206	2 US-08-599-491-32	Sequence 32, Appl
42	29	20.7	206	2 US-08-756-386-32	Sequence 32, Appl
43	29	20.7	206	2 US-08-823-516-27	Sequence 27, Appl
44	29	20.7	206	3 US-08-682-853A-32	Sequence 32, Appl
45	29	20.7	206	3 US-08-759-038-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-08-861-774E-83/c
Sequence 83, Application US/08861774E

Patent No. 6297007

GENERAL INFORMATION:
APPLICANT: Waters, Barbara

APPLICANT: Miao, Vivian

APPLICANT: Ho, Yap

APPLICANT: Tong, Seow

TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR

FILE REFERENCE: 9993-006

CURRENT FILING DATE: 1997-05-22

NUMBER OF SEQ ID NOS: 94

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 83

LENGTH: 1222

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Clone ps7

US-08-861-774E-83

Query Match

Best local similarity 25.1%; Score 35.2; DB 4; Length 1222;

Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 1176 GCATGCTCCCGCGCGCCGATGTCATGAGG 50

US-09-182-145-51

Patent No. 6387657

GENERAL INFORMATION:

APPLICANT: Bolstein, David A.

APPLICANT: Cohen, Robert

APPLICANT: Goddard, Audrey

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Lawrence, David A.

APPLICANT: Levine, Arnold J.

APPLICANT: Pennica, Diane

APPLICANT: Roy, Margaret Ann

APPLICANT: Wood, William I.

TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: P1176R2

```
; CURRENT APPLICATION NUMBER: US/09/182,145B
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: US 60/063,704
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: US 60/073,612
; EARLIER FILING DATE: 1998-02-04
; EARLIER APPLICATION NUMBER: US 60/081,695
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 156
; SEQ ID NO 51
; LENGTH: 163
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1-163
; OTHER INFORMATION: Sequence is synthesized
; Patent No. 6387657
US-09-182-145-51
```

```
Query Match          24.4%; Score 34.2; DB 4; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.0038;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 49
         |||||||
Db       41 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 79
```

```
RESULT 3
US-09-178-109-3/c
; Sequence 3, Application US/09178109
; Patent No. 6395477
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dilks, Daniel W.
; APPLICANT: Chang Ling, Hui-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; TITLE OF INVENTION: Polypeptides and Uses Therefor
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/09/178,109
; CURRENT FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: human
US-09-178-109-3
```

```
Query Match          24.3%; Score 34; DB 4; Length 2064;
Best Local Similarity 88.1%; Pred. No. 0.009;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 52
         |||||||
Db       2015 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 1974
```

```
RESULT 4
US-09-178-109-1/c
; Sequence 1, Application US/09178109
; Patent No. 6395477
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dilks, Daniel W.
; APPLICANT: Chang Ling, Hui-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; TITLE OF INVENTION: Polypeptides and Uses Therefor
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/09/178,109
```

```
; CURRENT FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2121
; TYPE: DNA
; ORGANISM: human
US-09-178-109-1
```

```
Query Match          24.3%; Score 34; DB 4; Length 2121;
Best Local Similarity 88.1%; Pred. No. 0.0091;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 52
         |||||||
Db       2072 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 2031
```

```
RESULT 5
US-08-991-789A-175
```

```
; Sequence 175, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
```

```
; NUMBER OF SEQUENCES: 292
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seed IP Law Group
```

```
; STREET: 701 Fifth Avenue, Suite 6300
```

```
; CITY: Seattle
```

```
; STATE: Washington
```

```
; COUNTRY: USA
```

```
; ZIP: 98104-7092
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
```

```
; FILING DATE: 11-Dec-1997
```

```
; CLASSIFICATION: <unknown>
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Potler, Jane E. R.
```

```
; REGISTRATION NUMBER: 33,332
```

```
; REFERENCE/DOCKET NUMBER: 210121.419C3
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
```

```
; TELEFAX: (206) 682-6031
```

```
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 175:
US-08-991-789A-175
```

```
Query Match          24.1%; Score 33.8; DB 4; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.0071;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCAACTAAACCTTG 64
         |||||||
Db       25 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCAACTAAACCTTG 78
```

```
RESULT 6
US-09-062-451-175
; Sequence 175, Application US/09062451
```

Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-175

Query Match 24.1% Score 33.8; DB 4; Length 484;
Best Local Similarity 75.9% Pred. No. 0.0071;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

DB 11 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCACTAACCCCTTG 64
25 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCACTAACCCCTTG 78

RESULT 7
US-09-598-326-175
Sequence 175, Application US/09598326
Patent No. 6423496
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 247
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group PLLC
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326

FILING DATE: 20-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 175:
US-09-598-326-175

Query Match 24.1% Score 33.8; DB 4; Length 484;
Best Local Similarity 75.9% Pred. No. 0.0071;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

DB 11 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCACTAACCCCTTG 64
25 GCATGCTCCCGCCGCGCATGCGCGGATGTCATGAGCACTAACCCCTTG 78

RESULT 8
US-09-004-838-93/c
Sequence 93, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
CONFERRING PEST RESISTANCE IN PLANTS
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-078810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 5506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE: -
NAME/KEY: -

LOCATION: 1..5506
OTHER INFORMATION: /note= "R32D"
US-09-004-838-93

Query Match 23.9%; Score 33.4; DB 4; Length 5506;
Best Local Similarity 97.1%; Pred. No. 0.019;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCGATGCGGATGTCGA 45
|||||
DB 5468 GCATGCTCCGCGCCGATGCGGATGTCGA 5434

RESULT 9
US-09-182-145-52/C
Sequence 52, Application US/09182145B
Patent No. 6387657
GENERAL INFORMATION:
APPLICANT: Botstein, David A.
APPLICANT: Cohen, Robert
APPLICANT: Goddard, Audrey
APPLICANT: Gueney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Lawrence, David A.
APPLICANT: Levine, Arnold J.
APPLICANT: Pennica, Diane
APPLICANT: Roy, Margaret Ann
APPLICANT: Wood, William I.
TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: P1176R2
CURRENT APPLICATION NUMBER: US/09/182,145B
EARLIER FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: US 60/063,704
EARLIER FILING DATE: 1997-10-29
EARLIER APPLICATION NUMBER: US 60/073,612
EARLIER FILING DATE: 1998-02-04
EARLIER APPLICATION NUMBER: US 60/081,695
EARLIER FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 156
SEQ ID NO 52
LENGTH: 163
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1-163
OTHER INFORMATION: Sequence is synthesized
Patent No. 6387657
US-09-182-145-52

Query Match 23.3%; Score 32.6; DB 4; Length 163;
Best Local Similarity 89.7%; Pred. No. 0.014;
Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCGATGCGGATGTCGATGAG 49
|||||
DB 123 GCATGCTCCGCGCCGATGCGGATGTCGATGAG 85

RESULT 10
US-09-214-095D-97
Sequence 97, Application US/09214095D
Patent No. 6280987
GENERAL INFORMATION:
APPLICANT: Landry, Donald
TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
FILE REFERENCE: 51400-A-PCT-US
CURRENT APPLICATION NUMBER: US/09/214,095D
CURRENT FILING DATE: 1999-07-19
NUMBER OF SEQ ID NOS: 121
SOFTWARE: PatentIn version 3.0
SEQ ID NO 97
LENGTH: 419

TYPE: DNA
ORGANISM: Murine
US-09-214-095D-97

Query Match 22.1%; Score 31; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCGATGCGGATGTCGAT 41
|||||
DB 18 GCATGCTCCGCGCCGATGCGGATGTCGAT 48

RESULT 11
US-09-214-095D-111
Sequence 111, Application US/09214095D
Patent No. 6280987
GENERAL INFORMATION:
APPLICANT: Landry, Donald
TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
FILE REFERENCE: 51400-A-PCT-US
CURRENT APPLICATION NUMBER: US/09/214,095D
CURRENT FILING DATE: 1999-07-19
NUMBER OF SEQ ID NOS: 121
SOFTWARE: PatentIn version 3.0
SEQ ID NO 111
LENGTH: 420
TYPE: DNA
ORGANISM: Murine
FEATURE:
NAME/KEY: V-segment
LOCATION: (1)..(403)
OTHER INFORMATION: n at any position represents any nucleotide including c,g,t,a,
US-09-214-095D-111

Query Match 22.1%; Score 31; DB 4; Length 420;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCGATGCGGATGTCGAT 41
|||||
DB 17 GCATGCTCCGCGCCGATGCGGATGTCGAT 47

RESULT 12
US-08-991-789A-185
Sequence 185, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Rudakia, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESS: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.

REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-991-789A-185

Query Match 22.1%; Score 31; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGCGCGGATT 41
11 |||||
DB 12 GCATGCTCCGGCCGCGCATGCGCGGATT 42

RESULT 13
US-09-062-451-185
Sequence 185, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
ATTORNEY/AGENT INFORMATION:
NAME: Smith, John M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David U.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-185

Query Match 22.1%; Score 31; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGCGCGGATT 41
11 |||||
DB 12 GCATGCTCCGGCCGCGCATGCGCGGATT 42

RESULT 14
US-09-598-326-185
Sequence 185, Application US/09598326
Patent No. 6423496
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
ATTORNEY/AGENT INFORMATION:
NAME: Smith, John M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 247
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group PLLC
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326
FILING DATE: 20-Jun-2000
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-598-326-185

Query Match 22.1%; Score 31; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGCGCGGATT 41
11 |||||
DB 12 GCATGCTCCGGCCGCGCATGCGCGGATT 42

RESULT 15
US-08-991-789A-167
Sequence 167, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
ATTORNEY/AGENT INFORMATION:
NAME: Smith, John M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
MEDIUM TYPE: floppy disk

Wed May 28 09:17:34 2003

us-09-743-533-14.rni

Page 6

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 167:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 167:
US-08-991-789A-167

22.0%; Score 30.8; DB 4; Length 500;
Query Match Best Local Similarity 94.1%; Pred. No. 0.082;
Matches 32; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Search completed: May 25, 2003, 20:18:29
Job time : 17.2026 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:57:50 : Search time 44.2464 Seconds
(without alignments)
4178.078 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
Sequence: 1 atagaatacagatgctcc.....gagctcgaatgcctcata 140

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA: *
1: /cgn2_6/pdata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/pdata/2/pubpna/PC7_NEW_PUB.seq: *
3: /cgn2_6/pdata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/pdata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/pdata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/pdata/2/pubpna/PC7US_PUBCOMB.seq: *
7: /cgn2_6/pdata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/pdata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/pdata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/pdata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/pdata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/pdata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/pdata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/pdata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.4	25.3	1630	9 US-10-098-841-1	Sequence 1, Appl
2	35.2	25.1	1222	10 US-09-924-256A-83	Sequence 83, Appl
3	34.4	24.6	3186	9 US-10-098-841-2	Sequence 2, Appl
4	34.2	24.4	163	9 US-10-112-267-51	Sequence 51, Appl
5	34	24.3	2064	12 US-10-062-879-3	Sequence 3, Appl
6	34	24.3	2121	12 US-10-062-879-1	Sequence 1, Appl
7	33.8	24.1	484	9 US-09-924-400-175	Sequence 175, App
8	33.8	24.1	484	10 US-09-810-936-175	Sequence 175, App
9	33.8	24.1	484	10 US-09-429-755-175	Sequence 175, App
10	33.2	23.7	711	10 US-09-772-134B-80	Sequence 80, Appl
11	32.6	23.3	163	9 US-10-112-267-52	Sequence 52, Appl
12	31.4	22.4	669	9 US-10-043-875-841	Sequence 841, App
13	31.4	22.4	713	10 US-09-772-134B-82	Sequence 82, Appl
14	31.4	22.4	831	9 US-09-894-916-6	Sequence 6, Appl
15	31	22.1	215	10 US-09-778-320-250	Sequence 250, App
16	31	22.1	215	10 US-09-910-689-250	Sequence 250, App
17	31	22.1	215	12 US-10-010-742-250	Sequence 97, Appl
18	31	22.1	419	9 US-09-940-727B-97	Sequence 111, App
19	31	22.1	420	9 US-09-940-727B-111	

20	31	22.1	581	10 US-09-728-952-25	Sequence 25, Appl
21	31	22.1	597	9 US-09-924-400-185	Sequence 185, App
22	31	22.1	597	10 US-09-810-936-185	Sequence 185, App
23	31	22.1	597	10 US-09-429-755-185	Sequence 185, App
24	31	22.1	1680	9 US-09-298-523B-73	Sequence 73, Appl
25	31	22.1	1766	9 US-09-298-523B-74	Sequence 38, Appl
26	31	22.1	2714	10 US-09-835-996A-38	Sequence 167, App
27	30.8	22.0	500	10 US-09-924-400-167	Sequence 167, App
28	30.8	22.0	500	10 US-09-810-936-167	Sequence 167, App
29	30.8	22.0	500	10 US-09-429-755-167	Sequence 180, App
30	30.4	21.7	488	10 US-09-924-400-180	Sequence 180, App
31	30.4	21.7	488	10 US-09-810-936-180	Sequence 180, App
32	30.4	21.7	488	10 US-09-429-755-180	Sequence 182, App
33	30.4	21.7	507	10 US-09-924-400-182	Sequence 182, App
34	30.4	21.7	507	10 US-09-810-936-182	Sequence 182, App
35	30.4	21.7	507	10 US-09-429-755-182	Sequence 91, Appl
36	30	21.4	360	9 US-09-940-727B-91	Sequence 24, Appl
37	30	21.4	17458	9 US-10-055-001A-25	Sequence 26, Appl
38	30	21.4	17476	9 US-10-055-001A-26	Sequence 26, Appl
39	30	21.4	17681	9 US-10-055-001A-26	Sequence 26, Appl
40	30	21.4	17682	9 US-10-055-001A-23	Sequence 13, Appl
41	29.8	21.3	5829	9 US-10-012-896-473	Sequence 473, App
42	29.8	21.3	5829	9 US-09-895-793-473	Sequence 472, App
43	29.8	21.3	5829	9 US-09-895-814-473	Sequence 473, App
44	29.8	21.3	5829	9 US-10-010-940-473	
45	29.8	21.3	5829	9 US-10-010-940-473	

ALIGNMENTS

RESULT 1
US-10-098-841-1
Sequence 1, Application US/10098841
Publication No. US20020197679A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungding
APPLICANT: Wang, JianRui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Qian, Xiaohong B.
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/10/098, 841
CURRENT FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598, 042
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/552, 317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488, 725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 331
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 1
LENGTH: 1630
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (261)..(1052)
US-10-098-841-1

Query Match 25.3%; Score 35.4; DB 9; Length 1630;
Best Local Similarity 79.2%; Pred. No. 0.0038;
Matches 42; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 11 GCATGCTCCCGCCGCGCATGGCGGATGTCATGACACTAACCCTT 63
10 GCATGCTCCCGCCGCGCATGGCGGATTTCCCGGGTTCGACGATTCGTT 62

RESULT 2
US-09-924-256A-83/c
Sequence 83, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yip
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 83
LENGTH: 1222
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Clone ps7
US-09-924-256A-83

Query Match 25.1%; Score 35.2; DB 10; Length 1222;
Best Local Similarity 92.5%; Pred. No. 0.0043;
Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 11 GCATGCTCCCGCCGCGCATGGCGGATTTGTATGAGG 50
1176 GCATGCTCCCGCCGCGCATGGCGGATTTGTATGAGG 1137

RESULT 3
US-10-098-841-2
Sequence 2, Application US/10098841
Publication No. US20020197679A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yuning
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-Hong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Wehtman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Qian, Xiaohong B.
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/10/098,841
CURRENT FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598,042
PRIOR FILING DATE: 2000-06-20

PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 331
SOFTWARE: pf_fl_genes Version 1.0
SEQ ID NO 2
LENGTH: 3186
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (22)..(1848)
US-10-098-841-2

Query Match 24.6%; Score 34.4; DB 9; Length 3186;
Best Local Similarity 86.4%; Pred. No. 0.01;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 11 GCATGCTCCCGCCGCGCATGGCGGATTTGATGAGGAC 54
20 GCATGCTCCCGCCGCGCATGGCGGATTTGCTTCATGCAC 63

RESULT 4
US-10-112-267-51
Sequence 51, Application US/10112267
Publication No. US2003006678A1
GENERAL INFORMATION:
APPLICANT: Botstein, David A.
APPLICANT: Cohen, Robert
APPLICANT: Goddard, Audrey
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Lawrence, David A.
APPLICANT: Levine, Arnold J.
APPLICANT: Pennica, Diane
APPLICANT: Roy, Margaret Ann
APPLICANT: Wood, William I.
TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: P1176R2
CURRENT APPLICATION NUMBER: US/10/112,267
PRIOR FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/182,145B
PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/063,704
PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/073,612
PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/081,695
PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 156
SEQ ID NO 51
LENGTH: 163
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1-163
OTHER INFORMATION: Sequence is synthesized
US-10-112-267-51

Query Match 24.4%; Score 34.2; DB 9; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.0063;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 11 GCATGCTCCCGCCGCGCATGGCGGATTTGATGAGG 49
41 GCATGCTCCCGCCGCGCATGGCGGATTTATCACTAG 79

RESULT 5
US-10-062-879-3/c


```
; Sequence 3, Application US/10062879
; Patent No. US20020127649A1
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dिल्s, Daniel W.
; APPLICANT: Chang Ling, Hual-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/10/062,879
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/178,109
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: human
US-10-062-879-3
```

```
Query Match 24.3%; Score 34; DB 12; Length 2064;
Best Local Similarity 88.1%; Pred. No. 0.013;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 11 GCATGCTCCCGCGCCGATGCGCGGATTTGTCATGAGCA 52
Db 2015 GCATGCTCCCGCGCCGATGCGCGGATTTTACAAAGCA 1974
```

RESULT 6

```
; Sequence 1, Application US/10062879
; Patent No. US20020127649A1
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dिल्s, Daniel W.
; APPLICANT: Chang Ling, Hual-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/10/062,879
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/178,109
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2121
; TYPE: DNA
; ORGANISM: human
US-10-062-879-1
```

```
Query Match 24.3%; Score 34; DB 12; Length 2121;
Best Local Similarity 88.1%; Pred. No. 0.013;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 11 GCATGCTCCCGCGCCGATGCGCGGATTTGTCATGAGCA 52
Db 2072 GCATGCTCCCGCGCCGATGCGCGGATTTTACAAAGCA 2031
```

RESULT 7

```
; Sequence 175, Application US/09924400
; Patent No. US20020165371A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Lynda E.
```

```
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; APPLICANT: Li, Samuel X.
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.419C12
; CURRENT APPLICATION NUMBER: US/09/924,400
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 175
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; OTHER INFORMATION: n = A,T,C or G
US-09-924-400-175
```

```
Query Match 24.1%; Score 33.8; DB 9; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.011;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY 11 GCATGCTCCCGCGCCGATGCGCGGATTTGTCATGAGCACTAAACCTTG 64
Db 25 GCATGNTCCCGCGCCGATGCGCGGATTCGGTGATGCTCTCAGGCTTG 78
```

RESULT 8

```
; Sequence 175, Application US/09810936
; Patent No. US20020068285A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Linda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.419C11
; CURRENT APPLICATION NUMBER: US/09/810,936
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(484)
; OTHER INFORMATION: n = A,T,C or G
US-09-810-936-175
```

```
Query Match 24.1%; Score 33.8; DB 10; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.011;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY 11 GCATGCTCCCGCGCCGATGCGCGGATTTGTCATGAGCACTAAACCTTG 64
Db 25 GCATGNTCCCGCGCCGATGCGCGGATTCGGTGATGCTCTCAGGCTTG 78
```

Query Match	23.7%	Score 33.2	DB 10	Length 711
Best Local Similarity	92.1%	Pred. No. 0.02		
Matches 35, Conservative	0	Mismatches 3	Indels 0	Gaps 0
QY	11	GCATGCTCCCGCGCCATGCGCGCGGCGGATGTATGCA	48	

```

? RESULT 12
? US-10-043-875-841/c
? : Sequence 841. Application US/10043875
? : Publication No. US20030054339A1
? : GENERAL INFORMATION:
? : APPLICANT: De Smet, Koenraad
? : APPLICANT: Stuyver, Lieven
? : TITLE OF INVENTION: Method for Detection of Drug-Induced Mutations in the HIV Reverse
? : TITLE OF INVENTION: Transcriptase Gene
? : FILE REFERENCE: 11362-0033-NU501 (INNS:033)
? : CURRENT APPLICATION NUMBER: US/10/043.875
? : CURRENT FILING DATE: 2002-04-03
? : PRIOR APPLICATION NUMBER: 60/286,102
? : PRIOR FILING DATE: 2001-04-24
? : PRIOR APPLICATION NUMBER: EP 01870085.6
? : PRIOR FILING DATE: 2001-04-20
? : PRIOR APPLICATION NUMBER: EP 01870005.4
? : PRIOR FILING DATE: 2001-01-11
? : NUMBER OF SEQ ID NOS: 884
? : SOFTWARE: PatentIn version 3.1
? : SEQ ID NO 841
? : LENGTH: 669
? : TYPE: DNA

```

```

: ORGANISM: Human immunodeficiency virus
US-10-043-875-841

Query Match
Best Local Similarity 22.4%; Score 31.4; DB 9; Length 669;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATTGT 43
    |||||||
DB 626 GCATGCTCCGGCCGCGCATGGCGGGGATTAT 594

RESULT 13
US-09-772-134B-82
: Sequence 82, Application US/09772134B
: Patent No. US2002014310A1
: GENERAL INFORMATION:
: APPLICANT: Southern Illinois University
: APPLICANT: Lightfoot, David
: APPLICANT: Meksem, Khalid
: TITLE OF INVENTION: ISOLATED POLYNUCLEOTIDES AND POLYPEPTIDES RELATING TO LOCI UNDERL
: TITLE OF INVENTION: RESISTANCE TO SOYBEAN CYST NEMATODE AND SOYBEAN SUDDEN DEATH SYN
: FILE REFERENCE: 1268/4/2
: CURRENT APPLICATION NUMBER: US/09/772.134B
: PRIOR FILING DATE: 2001-01-29
: PRIOR APPLICATION NUMBER: 60/178,811
: NUMBER OF SEQ ID NOS: 122
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 82
: LENGTH: 713
: TYPE: DNA
: ORGANISM: soybean
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)..(713)
: OTHER INFORMATION: n is an undetermined nucleotide (dntp, dtrp, or dtrp)
US-09-772-134B-82

Query Match
Best Local Similarity 22.4%; Score 31.4; DB 10; Length 713;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATTGT 43
    |||||||
DB 46 GCATGCTCCGGCCGCGCATGGCGGGGATTAT 78

RESULT 14
US-09-894-916-6/C
: Sequence 6, Application US/09894916
: Publication No. US20030032089A1
: GENERAL INFORMATION:
: APPLICANT: Hunter, Fiona F.
: APPLICANT: Bidochka, Michael J.
: TITLE OF INVENTION: Isolated Cocoon Silk Protein From Simulium vittatum And Nucleic A
: TITLE OF INVENTION: Encoding Such Protein
: FILE REFERENCE: 1468-001A
: CURRENT APPLICATION NUMBER: US/09/894.916
: CURRENT FILING DATE: 2001-06-28
: PRIOR APPLICATION NUMBER: 60/214992
: PRIOR FILING DATE: 2000-06-29
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6
: LENGTH: 831
: TYPE: DNA
: ORGANISM: Simulium vittatum
US-09-894-916-6

Query Match
Best Local Similarity 22.4%; Score 31.4; DB 9; Length 831;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATTGT 43
    |||||||
DB 818 GCATGCTCCGGCCGCGCATGGCGGGGATTAT 786

RESULT 15
US-09-778-320-250
: Sequence 250, Application US/09778320
: Patent No. US20010034052A1
: GENERAL INFORMATION:
: APPLICANT: Dillon, David C.
: APPLICANT: Day, Craig H.
: APPLICANT: Jiang, Yugu
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Mitcham, Jennifer
: APPLICANT: Wang, TongTong
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
: FILE REFERENCE: 210121.491C5
: CURRENT APPLICATION NUMBER: US/09/778.320
: CURRENT FILING DATE: 2001-02-06
: NUMBER OF SEQ ID NOS: 301
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 250
: LENGTH: 215
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)..(215)
: OTHER INFORMATION: n=A,T,C or G
US-09-778-320-250

Query Match
Best Local Similarity 22.1%; Score 31; DB 10; Length 215;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATT 41
    |||||||
DB 20 GCATGCTCCGGCCGCGCATGGCGGGGATT 50

Search completed: May 25, 2003, 20:21:58
Job time : 45.2464 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:52:25 ; Search time 487.844 Seconds
(without alignments)
4647.728 Million cell updates/sec

Title: US-09-743-533-14
Perfect score: 140
Sequence: 1 atagaatacagcatcgtcc.....gagctcgaattcgccctata 140

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_estlrc:*
9: gb_estl1:*
10: gb_estl2:*
11: gb_estl3:*
12: gb_estl4:*
13: gb_estl5:*
14: gb_estl6:*
15: em_estlun:*
16: em_estlro:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.4	27.4	521	9 AA255342	AA255342 T3598 Bio
2	32.8	23.4	402	9 AJ251073	AJ251073 AJ251073
3	32.4	23.1	686	13 BM579445	BM579445 170006872
4	32.4	23.1	733	13 BI273615	BI273615 A1241_1e
5	32.2	23.0	431	17 AO058015	AO058015 CIT-HSP-2
6	32.2	23.0	512	17 AZ366183	AZ366183 IM0115K24

C 7	32.2	23.0	559	13 BI262739	BI262739 EST003 Su
C 8	32.2	23.0	617	13 BI262738	BI262738 EST002 Su
C 9	32.2	23.0	729	13 BI262743	BI262743 EST008 Su
C 10	32.2	22.9	373	17 B30807	B30807 HS-1003-A2-
C 11	32.2	22.9	417	14 BQ094068	BQ094068 040802-A2-
C 12	32.2	22.9	442	17 AC015890	AC015890 CIT-HSP-2
C 13	31.8	22.7	237	14 BP018827	BP018827 BP018827
C 14	31.8	22.7	497	17 A1133472	A1133472 HA2082 Hu
C 15	31.8	22.7	1222	9 AB021948	AB021948 AB021948
C 16	31.6	22.6	362	17 B16177	B16177 347C21.TP C
C 17	31.4	22.4	446	17 B36957	B36957 HS-1048-B2-
C 18	31	22.1	150	9 AA012772	AA012772 RPU2008GC
C 19	31	22.1	185	9 A1738079	A1738079 ESTRGT02
C 20	31	22.1	187	9 AB029050	AB029050 AB029050
C 21	31	22.1	211	10 AM888331	AM888331 E45 Rat d
C 22	31	22.1	252	9 A1738086	A1738086 ESTRGT09
C 23	31	22.1	258	13 BM139628	BM139628 Gm-35-10H
C 24	31	22.1	277	9 AA052854	AA052854 T3513 Bio
C 25	31	22.1	319	9 AA585085	AA585085 T3655 Bio
C 26	31	22.1	320	9 AA585076	AA585076 T3645 Bio
C 27	31	22.1	335	9 AA585077	AA585077 T3646 Bio
C 28	31	22.1	365	9 AA052853	AA052853 T3507 Bio
C 29	31	22.1	393	12 BG006643	BG006643 RC2-GN026
C 30	31	22.1	408	13 BI203089	BI203089 PC89R1 Pe
C 31	31	22.1	438	14 AM888346	AM888346 MEG4 Rat
C 32	31	22.1	474	14 B0094058	B0094058 040802-17
C 33	31	22.1	581	13 BM139960	BM139960 Gm-R2 Soy
C 34	31	22.1	601	10 AM224940	AM224940 EST006 P1
C 35	31	22.1	629	10 AV735727	AV735727 AV735727
C 36	31	22.1	637	11 AF092132	AF092132 Homo sapi
C 37	31	22.1	190	9 AA052864	AA052864 T3541 Bio
C 38	30.8	22.0	377	9 AT006477	AT006477 AF006477
C 39	30.8	22.0	480	10 BE411794	BE411794 ISC009 ID0
C 40	30.6	21.9	468	13 BM359493	BM359493 GA_Ea002
C 41	30.6	21.9	565	9 A1728509	A1728509 BNLG1109
C 42	30.6	21.9	583	12 BG443394	BG443394 GA_Ea002
C 43	30.6	21.9	593	9 A1728966	A1728966 BNLG1121
C 44	30.6	21.9	639	17 AO110909	AO110909 CIT-HSP-2
C 45	30.6	21.9	17	AO110909	CIT-HSP-2

ALIGNMENTS

RESULT 1
AA255342
LOCUS
DEFINITION
AA255342
brucei CDNA 5', mRNA sequence.
ACCESSION
AA255342
VERSION
AA255342.1 GI:188945
KEYWORDS
EST.
SOURCE
Trypanosoma brucei brucei.
ORGANISM
Trypanosoma brucei brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
1 (bases 1 to 521)
Osanya, A., Pelle, R. and Murphy, N.B.
Identification and characterisation of differentially expressed
genes of Trypanosoma brucei brucei
Unpublished (1997)
JOURNAL
Contact: Osanya A
Unit 3
International Livestock Research Institute
Box 30709, Nairobi, Kenya
Tel: 254 2 630 743
Fax: 254 2 631 499
Email: a.osanya@cgiar.org
Seq primer: M13 primer.
Location/Qualifiers
1..521
/organism="Trypanosoma brucei brucei"
/strain="Clone A4"

FEATURES

Source

/db.xref="taxon:5702"
/clone.lib="Bloodstream form of serodeme ITat1.1"
/note="CDNAs were generated from poly (A+) enriched RNA prepared from different developmental stages of T.b. brucei by reverse transcription followed by PCR amplification using mini-exon and oligo(dT) primers. The cDNA generated were utilized in RABES-PCR coupled with differential hybridisations to identify differentially expressed mRNA transcripts. The products which showed to be differentially expressed were cloned pGEM-T vector. These differentially expressed mRNA transcripts were (are being) sequenced to generate differentially expressed sequence tags."

BASE COUNT 151 a 140 c 117 g 113 t

ORIGIN

Query Match 27.4%; Score 38.4; DB 9; Length 521;
Best Local Similarity 62.5%; Pred. No. 0.38;
Matches 60; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 11 GCATGCTCCCGCGCCGATGTCATGAGCCACTAACCCCTGCGAGCG 70
|||||
Db 142 GCATGCTCCCGCGCCGATGTCGCGCGGATTCGCGCCACCATTTATTCGCACT 201
|||||

OY 71 CCCCCACAGACTTCATGATGAGTATGCTCTA 106
|||||
Db 202 TCCCTTCATGCACTCAAGCACAGATGTTAGTGGGATA 237
|||||

RESULT 2
AJ251073 402 bp mRNA linear EST 29-NOV-1999
LOCUS AJ251073 Rattus norvegicus library (Ma M2) Rattus norvegicus cDNA
DEFINITION clone #1, mRNA sequence.
ACCESSION AJ251073
VERSION AJ251073.1 GI:6470363
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 402)
AUTHORS Ma,Z.M., Yang,Y., Xie,Y.Z. and Fan,M.
TITLE The DB-PCR studies of the candidate genes of rat brain under system hypoxia
JOURNAL Unpublished (1999)
COMMENT Contact: Ma M2
Department of Neurobiology Sciences
Institute of Basic Medical Sciences
No.27, Tai Pin Road, Beijing, 100850, China.
FEATURES
SOURCE
1. 402
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone.lib="Rattus norvegicus library (Ma M2)"
/note="Organ: brain; Vector: pGEM-T; with treatment by endostatin, the differential expressed cDNAs were cloned into the pGEM-T vector. The 3' anchor primer is AAGCTTTTCTTTT, and the 5' arbitrary primer is 10-mer oligo. Individual clone was sequenced with T7 promoter primer."

BASE COUNT 113 a 89 c 87 g 112 t 1 others

ORIGIN

Query Match 23.4%; Score 32.8; DB 9; Length 402;
Best Local Similarity 94.4%; Pred. No. 16;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 11 GCATGCTCCCGCGCCGATGTCATGATTCAT 46
|||||
Db 58 GCATGCTCCCGCGCCGATGTCATGATTCAT 93
|||||

RESULT 3
BM579445 686 bp mRNA linear EST 22-FEB-2002
LOCUS BM579445
DEFINITION 17000687228519 A.gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
19600449682280 5', mRNA sequence.
ACCESSION BM579445
VERSION BM579445.1 GI:18867912
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
REFERENCE 1 (bases 1 to 686)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
TITLE R., Collins,F.H., Venter,J.C., and Hoffman,S.L.
JOURNAL Celeris Anopheles gambiae EST project
COMMENT Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AZX row: C column: 02
Seq primer: M13 Reverse.

FEATURES
source
1. 686
/organism="Anopheles gambiae"
/strain="RSP-St (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone.lib="A.gam.ad.cDNA.blood1"
/dev_stage="Adult"
/lab_host="DHI0b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"

BASE COUNT 141 a 157 c 199 g 189 t

ORIGIN

Query Match 23.1%; Score 32.4; DB 13; Length 686;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 12 CATGCTCCCGCGCCGATGTCATGAGCACTAACCTTGCAGCGTC 71
|||||
Db 281 CAGCGCTCGCGCGCTTGGCCGCGCTGCTTCATTTGCGAGCAAGAAC 340
|||||

OY 72 CCCCACAGACTTCATGATGAGATG 101
|||||
Db 341 GCCACACAGTTAATGCAATGTGTATGTG 370
|||||

RESULT 4
BI273615 733 bp mRNA linear EST 12-JUN-2002
LOCUS BI273615 Aplysia metacerebral cell cDNA library Aplysia californica
DEFINITION cDNA 5', mRNA sequence.
ACCESSION BI273615
VERSION BI273615.1 GI:18020448
KEYWORDS EST.
SOURCE California sea hare.
ORGANISM Aplysia californica
Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae; Aplysiidae; Aplysia.
REFERENCE 1 (bases 1 to 733)

Query Match	Best Local Similarity	Score	DB	Length
Matches	80: Conservative	23.0%; 56.7%;	DB 17; Pred. No. 25;	431;
		0: Mismatches	58; Indels	3; Gaps
				1;

BASE COUNT	112 a	100 c	117 g	102 t
ORIGIN	HindIII"			

QY	120	CGAGCTCGCAATTCGCCCTATA	140
Db	24	CGAGCTCGCAATTCGCCCTATA	4

RESULT 6	AZ366183/c	512 bp	DNA	linear	GSS 02-OCT-2000
LOCUS	AZ366183/c	512 bp	DNA	linear	GSS 02-OCT-2000
DEFINITION	clone UNGC1M0115K24 F, DNA sequence.				
ACCESSION	AZ366183				
VERSION	AZ366183.1	GI:10479883			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Mus.				
AUTHORS	1 (bases 1 to 512)				
	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.				
TITLE	Muscle whole genome scaffolding with paired end reads from 10kb				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0115 row: K column: 24 Seq primer: CGTGTAAACGACGCGCACAC Class: plasmid ends High quality sequence stop: 512. Location/Qualifiers 1. 512				

FEATURES	source
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/clone="UNGCM0115K24"	
/clone_lib="Mouse 10kb plasmid UNGCM library"	
/sex="Male"	
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"	
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adapter oligonucleotides were ligated to the blunt ends in high molar excess. The	

449 Chem/Biochem, Manhattan, KS 66506, USA
 Tel: 785 532 6964
 Fax: 785 532 7278
 Email: kanostek@ku.edu
 Seq primer: SP6

High quality sequence stop: 730.

FEATURES
 source
 Location/Qualifiers
 1..729

/organism="Manduca sexta"
 /db_xref="taxon:7130"
 /clone="008"
 /clone_lib="Subtracted larval fat body library,
 bacteria-induced genes"
 /tissue_type="Fat Body"
 /dev_stage="5th instar larva"
 /note="Vector: pGEM-T Vector; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 158 a 215 c 187 g 165 t 4 others
 ORIGIN

Query Match 23.08; Score 32.2; DB 13; Length 729;
 Best Local Similarity 91.9%; Pred. No. 32;
 Matches 34; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 GCATGCTCCGCGCCGATGCGGATTCGATG 47
 141 GCATGCTCCGCGCCGATGCGGATTCGATG 105

RESULT 10

B30807 373 bp DNA linear GSS 17-OCT-1997
 LOCUS HS-1003-A2-B01-MR.abi CIT Human Genomic Sperm Library C Homo

DEFINITION Sapiens genomic clone Plate=CT 497 Col=2 Row=C, DNA sequence.

ACCESSION B30807
 VERSION B30807.1 GI:2530176

KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 373)
 Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,

Tralcoff,R., Abdjian,C., Blanchard,A., West,A. and Hood,L.E.
 Construction of a Characterized Clone Resource for Genomic

Sequencing: Generation and Preliminary Analysis of 20,000 Sequence

Tagged Connectors
 Unpublished (1997)
 Contact: Mahairas GG, Zackrone KD, Hood L

University of Washington
 Seattle, WA 98195, USA

Tel: (206) 616-8744
 Fax: (206) 685-7301

Email: kzackrone@u.washington.edu
 Sequence Tagged Connector

Plate: CT 497 row: C column: 2
 Class: BAC ends

High quality sequence stop: 373.
 Location/Qualifiers

1..373
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=CT 497 Col=2 Row=C"
 /clone_lib="CIT Human Genomic Sperm Library C"

/sex="M"
 /note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"
 BASE COUNT 104 a 110 c 80 g 76 t 3 others

ORIGIN

Query Match 22.9%; Score 32; DB 17; Length 373;
 Best Local Similarity 79.2%; Pred. No. 27;
 Matches 38; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 93 TGCAGTATGCTAGGATCCGGTACGACCTCGAATTCGCCGATA 140
 16 TGCAGTATGCTAGGATCCGGTACGACCTCGAATTCGCCGATA 63

RESULT 11
 B0094068 417 bp mRNA linear EST 08-APR-2002
 LOCUS 040802_27 Rat male liver Rattus norvegicus cDNA 5', mRNA sequence.

DEFINITION B0094068
 ACCESSION B0094068.1 GI:20075285

KEYWORDS EST.
 SOURCE Norway rat.

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 417)
 Lee,Y. and Xu,C.

Expressed sequence tags from cDNA subtractive library following
 short interval successive partial hepatectomy in rat liver

regeneration
 Unpublished (2002)
 Contact: Lee Yuchang

Henan Bioengineering Key Lab
 Henan Normal Univ
 Jianshe Road, Xinxiang City, P.R.China

Tel: 0086373328084
 Fax: 0086373326524

Email: yuchang910@163.com
 Seq primer: M13 Forward.

location/Qualifiers
 1..417

/organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="Rat male liver"

/sex="male"
 /tissue_type="liver"

/note="Vector: pT73D-Pac (Pharmacia)-with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UTR-D01

library is a normalized Rat Cell Line R3327-5A library
 (mRNA) constructed in pT73 PAC vector according to the

procedure described by Bonaldo, Lennon & Soares
 (Normalization and Subtraction: Two Approaches to

Facilitate Gene Discovery. Genome Research 6: 791-806,
 1996). The oligonucleotide used to prime first strand

synthesis contained the sequence tag GGACGATGTC between
 the Not I cloning site and G118 stretch. The Rat Cell Line

R3327-5A was provided by Mary Hendrix of the University of
 Iowa."

BASE COUNT 66 a 124 c 97 g 130 t

ORIGIN

Query Match 22.9%; Score 32; DB 14; Length 417;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGCGCCGATGCGGATTCGATG 42
 19 GCATGCTCCGCGCCGATGCGGATTCGATG 50

RESULT 12
 A0015890 442 bp DNA linear GSS 09-JUN-1998

LOCUS CIT-HSP-2314E8.TF CIT-HSP Homo sapiens genomic clone 2314E8, DNA

DEFINITION A0015890
 ACCESSION A0015890

VERSION A0015890.1 GI:3194626
 KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.

AUTHORS

1 (bases 1 to 1222)

Toriyama, K. and Ohhashi, Y.

JOURNAL

EST of immature anther of rice

COMMENT

Unpublished (1999)
Contact: Toriyama K
Faculty of Agriculture, Laboratory of Plant Breeding
Tohoku University
Tsuetsunimori-Amamiyamachi 1-1, Aobaku, Sendai, Miyagi 981-8555,
Japan
Email: torikin@bios.tohoku.ac.jp.
Location/Qualifiers

FEATURES

1. 1222

/organism="Oryza sativa"

/strain="ms-bo-Taichung 65"

/db_xref="taxon:4530"

/clone="DD2R3-1"

/clone_id="Oryza sativa ms-bo-Taichung 65 immature
anther"

/tissue_type="anther"

/dev_stage="immature anther"

BASE COUNT

297 a 298 c 291 g 335 t 1 others

ORIGIN

Query Match 22.7%; Score 31.8; DB 9; Length 1222;

Best Local Similarity 94.3%; Pred. No. 52;

Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCCATGCGCGGATTTCA 45

DB 16 GCATGCTCCGCGCCCATGCGCGGATTTATTA 50

Search completed: May 25, 2003, 20:17:10
Job time : 491.844 secs

—

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	378.8	79.4	1420	8	HYHON117	X60037 H.vulgaris b
2	329.6	69.1	2806	8	S66938	S66938 C-hordein l
3	231.4	48.5	3925	8	AF280606	AF280606 Triticum
4	230.6	48.3	9248	8	SCAF000227	AF000227 Secale ce
5	229	47.0	1832	8	SCSECB1	X60295 S.cereale s
6	224.2	47.0	1342	8	SCSECB1	X60294 S.cereale s
7	219.6	46.0	3789	8	AF280605	AF280605 Triticum
8	216.6	45.4	2065	8	BYH0800CA	M65641 Hordeum vul
9	200.6	42.1	1858	8	AB055812	AB055812 Triticum
10	148.2	31.1	4971	8	AF201084	AF201084 Secale ce
11	140.4	29.4	927	8	AF175312	AF175312 Triticum
12	140.4	29.4	1178	8	AF234650	AF234650 Triticum
13	140.4	29.4	1397	8	WHTGLIGV	M16064 Wheat (T. ae
14	139.4	29.2	771	8	AF234642	AF234642 Triticum
15	139.4	29.2	1227	8	AF234644	AF234644 Triticum
16	135.2	28.3	5718	8	AF234647	AF234647 Triticum
17	134.6	28.2	730	8	ASO389679	AJ389679 Aegilops
18	134.6	28.2	757	8	TME389666	AJ389666 Triticum
19	134.6	28.2	777	8	TME389668	AJ389668 Triticum
20	134.6	28.2	778	8	TSP389674	AJ389674 Triticum
21	134.6	28.2	792	8	TGCO389669	AJ389669 Triticum
22	134.6	28.2	798	8	WHTGLIGP	M16060 Wheat (T. ae
23	134.6	28.2	800	8	TMA389671	AJ389671 Triticum
24	134.6	28.2	801	8	TMA389673	AJ389673 Triticum
25	134.6	28.2	801	8	TSP389676	AJ389676 Triticum
26	134.6	28.2	809	8	TMA389672	AJ389672 Triticum
27	134.6	28.2	850	8	TCO389670	AJ389670 Triticum
28	134.6	28.2	947	8	AF120267	AF120267 Triticum
29	134.6	28.2	956	8	AF144104	AF144104 Triticum
30	133.6	28.0	2086	8	WHTGGLN	M36599 Wheat genome
31	132	27.7	842	8	TGCGAGL	X53412 Wheat gene
32	131.8	27.6	750	8	TME389667	AJ389667 Triticum
33	131.8	27.6	762	8	TVA389678	AJ389678 Triticum
34	131.6	27.5	731	8	ASO389680	AJ389680 Aegilops
35	130.6	27.4	731	8	TSP389675	AJ389675 Triticum
36	129.6	27.2	746	8	TVA389677	AJ389677 Triticum
37	128.2	26.9	6443	8	AF234649	AF234649 Triticum
38	127.4	26.7	604	8	ASO389685	AJ389685 Aegilops
C 39	126.6	26.5	156533	2	AC117070	AC117070 Dictyoste
C 40	125.6	26.4	622	8	ASO389687	AJ389687 Aegilops
C 41	124.6	26.1	97683	2	AC116548	AC116548 Dictyoste
42	124.4	26.1	718	8	TPU389703	AJ389703 Triticum
43	124.4	26.1	886	8	TPU389704	AJ389704 Triticum
44	124.2	26.0	607	8	ASO389688	AJ389688 Aegilops
45	124.2	26.0	611	8	ASO389684	AJ389684 Aegilops

ALIGNMENTS

RESULT	1
Locus	HVHOR17
DEFINITION	HVHOR17 1420 bp DNA
ACCESSION	X60037 linear PLN-05-MAY-1995
VERSION	X60037.1 GI:19000
KEYWORDS	endosperm; seed storage protein.
SOURCE	Hordeum vulgare subsp. vulgare.
ORGANISM	Hordeum vulgare subsp. vulgare. Eukaryota; Viridiplantae; Streptophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poaceae; Triticeae; Hordeum 1 (bases 1 to 1420)
REFERENCE	Entwistle,J.
AUTHORS	Direct Submission
TITLE	

[illegible]

Query Match	69.1%;	Score 329.6;	DB 8;	Length 2806;
Best Local Similarity	85.1%;	Pred. No. 2.6e-56;		
Matches 400;	Conservative	0;	Mismatches 49;	Indels 21; Gaps 2;
QY	4	AGGCACCTAAACCCCTGGACGCAAGATGTGCAATCCACCAACAAATCATATGTGGCGAG	63	
Db	1144	AGGCACCTAAACCTAGACGCGCAAGATGTGCAATCCACCAACAAATCATATGTGCGACG	1203	
QY	64	CCATATCCACAAACCCATATCTACCGCAAAACATTTTCAATGTGACGACACCGTTTAC	123	
Db	1204	CCATATCCACAAACCCATATCTACCGCAACAAACATTTTCAATGTGACGACACCGTTTAC	1263	
QY	124	ACACCCCAACAATATTTTCCCTATCTACAGAGGAATTTGTTCCCATATCAAAATATACA	183	
Db	1264	ACACCCCAACAATATTTTCCCTATCTACAGAGGAATTTGTTCCCATATCAAAATATACA	1322	
QY	184	ACCCCTTACACACCAACACATCTCCCAACAACAACCTCTCTCGGGCC	243	
Db	1324	ACCCCTTACACACCAACACATCTCCCAACAACAACCTCTCTCGGGCC	1383	
QY	244	CANCAACCAATTCCTGGCAACCAACAACATTTTCCCGACGCCCAAGAACCAATTTCC	303	

Db 1384 CAACAACGATTCCTCCGCGACCAACAACGATTTCCCGAGCCCAACAACCAATTTCC 1443

QY 304 -----CAACACCATTTCCCGGCAACCAACAACCAATTTCCCGAGCCCAAGA 354

Db 1444 TACCAACCAACAACCAATTTCAACGACCAACAACAATATATATCCAGCAACCA 1503

QY 355 -----CCAAATTCACAATAATATTTTCAGACCAACCCCAACATCATACCTGTG 402

Db 1504 CAACATTCCTCCCAACCAACCAACCTTTTCCACAGCCCAACAACATTTCCCTG 1563

QY 403 CAACCTCAACGACGATTTCTCAACACCTCAACCAAGTCCCAACCAAC 452

Db 1564 CAACCAACAACCAATTTCCAGGCTCAACAACCAATTTCCCTGCAAC 1613

RESULT 3

AF280606 Locus AF280606 3925 bp DNA Linear PLN 02-OCT-2000

DEFINITION Triticum aestivum omega gliadin pseudogene, complete sequence.

ACCESSION AF280606

VERSION AF280606.1 GI:10444085

KEYWORDS Triticum aestivum.

SOURCE Triticum aestivum.

ORGANISM Triticum aestivum.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

AUTHORS 1 (bases 1 to 3925)

TITLE Hsia,C.C. and Anderson,O.D.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3925)

AUTHORS Hsia,C.C. and Anderson,O.D.

TITLE Direct Submission

JOURNAL Submitted (20-JUN-2000) Western Regional Research Center, Agricultural Research Service, 800 Buchanan Street, Albany, CA 94710, USA

FEATURES

source Location/Qualifiers

1..3925

/organism="Triticum aestivum"

/db_xref="taxon:4565"

/chromosome="1"

/clone="omega3"

/dev_stage="embryo"

2372..2377

2477..3550

/note="prolamine storage protein"

/pseudo

/codon_start=1

/product="omega gliadin"

3026..3028

/note="in-frame stop codon"

polya_signal 3626..3631

polya_signal 3683..3688

BASE COUNT 1327 a 995 c 559 g 1044 t

ORIGIN

Query Match 48.5%; Score 231.4; DB 8; Length 3925;

Best Local Similarity 74.5%; Pred. No. 1.2e-36;

Matches 327; Conservative 0; Mismatches 91; Indels 21; Gaps 2;

QY 40 CCACAACATCATATTCGCGCAGCATATCCAGAAAACCATATTCAGCAAAACCA 99

Db 2594 CAACAACCAACATTTCCAGAGATATTCACAAACACATTCATCATCATCAACCA 2653

QY 100 TTTCAGTCGACGACGATTCACACACCCCAACAATATTTCCCTATCTACAGAGAA 159

Db 2654 TATCATTTGCAACAACCATTTCTACACCCCAACAATAATTTCCCGACATTCACAA 2713

QY 160 TTGTTTCCCAATATCAATACAAACCCCTTACACCAACCAACCAATTCACCA 219

Db 2714 CCAATTTCCCAACCAACCAACCAACCCCTTACACCAACCAACCAATTCACCA 2773

QY 220 CCACAACACGCTTCTTCGCGCCCAACAACCAATTTCCCTGCAACAACAACCAATTT 279

Db 2774 CCACAACCAACATTTTCACAGCCCAACAACCAATTTCTTGGCAACCAACAACCAATTT 2833

QY 280 CCCCAGCCCAAGAACCAAT-----TCCCAACAACCATTCCTCGCAACCAACCA 330

Db 2834 CCCCAGACCCCAACATGCTTCTCTGTGCAACCAACAGCCATTTCCCGACCAACCCCA 2893

QY 331 CAACATTTCCCGCAGCCCAAGAACCAATTC-----AACAAATATATTTTCAG 378

Db 2894 CAACATTTCCCGCAGCCCAACCAATTCACATTCACCAACAACCAACAATAATTTCCCGAG 2953

QY 379 CAACCCCAACAATATACCTGTGCAACCTCAACAGCCATTTCTCTCAACAACCTCAACA 438

Db 2954 CCAACCAACAACCAATTCCTCTGTGCAACCAACCAACCAATTTCCCGCAACCAATTC 3013

QY 439 GTCCCCCAACAACGCTTCAT 457

Db 3014 CAGCAACCCCAATTAACCAT 3032

RESULT 4

SCAF000227

LOCUS SCAF000227 9248 bp DNA Linear PLN 31-MAY-1997

DEFINITION Secale cereale omega secalin gene, complete cds.

ACCESSION AF000227

VERSION AF000227.1 GI:2145024

KEYWORDS Secale cereale.

SOURCE Secale cereale.

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Secale.

REFERENCE 1 (bases 1 to 9248)

AUTHORS Clarke,B.C., Mukai,Y. and Appels,R.

TITLE The Sec-1 locus on the short arm of chromosome 1R of rye (Secale cereale)

JOURNAL Chromosoma 105 (5), 269-275 (1996)

MEDLINE 97094352

PUBMED 8939819

REFERENCE 2 (bases 1 to 9248)

AUTHORS Clarke,B. and Appels,R.

TITLE Sequence variation at the Sec-1 locus of rye

JOURNAL Plant Syst. Evol. (1997) In press

REFERENCE 3 (bases 1 to 9248)

AUTHORS Clarke,B. and Appels,R.

TITLE Direct Submission

JOURNAL Submitted (17-APR-1997) Plant Industry, C.S.I.R.O., Clunies Ross, Canberra, ACT 2601, Australia

FEATURES

source Location/Qualifiers

1..9248

/organism="Secale cereale"

/db_xref="taxon:4550"

/chromosome="1"

/map="1RS"

/tissue_type="endosperm"

1..9248

/note="there are 15 omega secalin genes at the Sec-1 locus; this sequence covers one complete unit of this multigene family"

<2751..>3825

/product="omega secalin"

<2751..>3825

/note="there are 15 omega secalin genes at the Sec-1 locus"

2752..3825

/function="seed storage protein"

/codon_start=1

/product="omega secalin"

/protein_id="AA058403.1"

/db_xref="GI:2145025"

/translation="MKTFLEFVLAHTMSITTRARLNPSEDELSPQGPVPEKESYPO

QY	184	ACCCCCCTTCAACGACACAAACAACTATTCCTGGGCC	243
Db	364	ACCCCTATTACGACAAACAACTATTCCTGGGCC	423
QY	244	CACAAACAATTCCTGGCAACACAACTATTCCTGGGCC	303
Db	424	CACAAACAATTCCTGGCAACACAACTATTCCTGGGCC	483
QY	304	CAACAAACA-----TTCCCTGGCAACACAACTATTCCTGGGCC	354
Db	484	CAGCAACACCAACAACTATTCCTGGCAACACAACTATTCCTGGGCC	543
QY	355	CCAATTCAACAATATTTCCAGCAACCCCAACAACTATTCCTGGGCC	414
Db	544	CAACCAACAACAATATTTCCAGCAACCCCAACAACTATTCCTGGGCC	603
QY	415	CCATTTCCCAACAACTATTCCTGGGCC	435
Db	604	CCATTTCCCAACAACTATTCCTGGGCC	624

```

FEATURES
source
Location/Qualifiers
1..3789
/organism="Triticum aestivum"
/db_xref="taxon:4565"
/chromosome="1"
/clone="OmegaF20b"
/dev_stage="embryo"
2359..2364
<2464..23441
/product="Omega gliadin storage protein"
2464..3306
/codon_start=1

```

protein_10="AAG1/702.1"	/db_xref="GI:10444084"				
/translation="MKTFLLFLVLLAMAMKITAARELNPSNKELOSPOOSFSYQOOPQ					
PQOYPOOPVPSOOPVPSOOPPPQOQPEOSQOPQOPQOPPTIOPQOPFPQOP					
PQOQFPQOPQOPQOPQOPQOPQOOSFFIQTQOPPPQOPQOPFPOLIPFOOSEQII					
ELQOPFIQOPQOPFPQOPQOPQOPQOPFIQVQQSFPQOOSQOOSQOPFAQPOLPP					
IGGG					
polyA_signal	3379	.3384			
polyA_signal	3436	.3441			
BASE COUNT	1249 a	906 c	601 g	1033 t	
ORIGIN					
Query Match	46.08;	Score 219.6;	DB 8;	Length 3789;	
Best Local Similarity	74.38;	Pred. No. 2,7e-34;			
Matches 330; Conservative	0;	Mismatches 84;	Indels 30;	Gaps 3;	

QY	25	CAAGAGTTGGCAATCACCAACAACATATCTATCTGCGCAGACCATATCTCACAAACCCATAT	84
Db	2566	CAATCATTTTCTTATATCAACAACAACCATTTTCCACAGCAGCATATCTCAACAACAACCATAT	2625
QY	85	CTACCGCAAAACCATTTCCAGTGCAGCAACGGTTTCACACACCCCAACATATTTCC	144
Db	2626	CCATTCACAGCAACCATATTCATGCAACCAACCATTTTCCACACCCCAACATATTTCC	2685
QY	145	TATCTACCAAGAGATTTGTTTCCCATATTCMAATATCCAAACCCCTTAACAACAACAACAA	204
Db	2686	GAGCAATTCACAAACAACCATTTTACCCAGCCCAACCAACGACCCCATATCAACAACAACAA	2745
QY	205	CCATTTT-----CCCAACAACCAACAACCTTTCTTGCGCCCAACAACCATTC	255
Db	2746	CCATTTCCCCGACGACCCCAACAACAACCAACACTTTTCTACACAACCCCAACAACCATTT	2805
QY	256	CCCTGGCAACCAACAACCATTTTCCGAGCCCAAGAACCAT-----TCCCCAA	306
Db	2806	CCCTGGCAACCAACAACCATTTTCCGAGCCCAACAACCTTTTCTTCCCTTCCAAACCAAA	2865
QY	307	CAACCATTTCCCTGGCAACAACAACAACATTTTCCGAGCCCAAGAACCATTTTC-----	361
Db	2866	CAGCATTTTCCCCGACCAACCCCAACAACCATTTTCCAGCCCACTACCATTTCCCCAG	2925
QY	362	-----AACAAATATTTTCCAGCAACCCCAACAACATATACCTGTGCAACCTCAACAG	414
Db	2926	CAATCAGAAACAATAATTTCCACAGCAACTCCACAAACATTTCCCTGTGCAACCGCAACAA	2985
QY	415	CCATTTTCTCTCAACAACCTCAACCA	438
Db	2986	CCATTTTCTCTCAACAACCTCAACCA	3009

RESULT	8
LOCUS	BLYHORDCA . 2065 bp DNA linear PLN 15-MAR-1996
DEFINITION	Hordeum vulgare C-hordein gene, complete cds.
ACCESSION	M36941
VERSION	M36941.1 GI:167062
KEYWORDS	alcohol-soluble prolamins; hordein; seed storage protein.
SOURCE	Hordeum vulgare subsp. vulgare.
ORGANISM	Hordeum vulgare subsp. vulgare Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum. 1 (bases 1 to 2065) Entwistle,J. Primary structure of a C-hordein gene from barley Carlsberg Res. Commun. 53 (4), 247-258 (1988)
AUTHORS	JOURNAL
TITLE	MEDLINE
PUBMED	3255313
FEATURES	Location/Qualifiers 1..2065
SOURCE	

```

sig_peptide
mat_peptide

/organism="Hordeum vulgare subsp. vulgare"
/strain="Bomi"
/db_xref="taxon:112509"
/tissue_type="immature endosperm"
429..1472
/codon_start=1
/translation_except(pos:909..911,aa:OTHER)
/product="C hordeain"
/protein_id="AA092333.1"
/db_xref="GI:893242"
/translation="MKITPLTFVLLAMNSITVTARQLNPSHOELQSPQDPFLKQOSTY
QDPYQDPYLPQDPFRPLPQDFPFLPQDITFPSSQDPNLDQDPFLDPDPDPDPDP
PQDPDPDPDPDPFRPQDPQIVPQDPQDPFPQDPQDPQDPQDPSPMQDPDFLPDP
XPLQACQDPFLPQDLFPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP
PQDLQDPQDPQDIIQDPQDPQDPFLQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP
QIITQDPFLPQDPFLPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP
AQQDQSDIALLG"
429..488
489..1469
/product="C hordeain"

```

[illegible]

OY 4 AGGCAACTAAACCTTTCAGCCAGCAGAGTGTGAATCACCACAAACATATATCTTGGCCAG 63
 Db 489 AGGCAAGTAAACCTTACCGACCAAGAGTGTGAATCACCACAAACATTTCTGAACAA 548
 OY 64 CCATATCCACAAAACCATATCTACCGCAAAAACCATTTCCAGTGGCAGCAACCTTTCAC 123
 Db 549 CAATCATATCTGCACAACCATATTCACAAACACATATCTACGGCAACCATTTCCC 608
 OY 124 ACACCCCAACATATTTTCCCTATCTACGACAGAGATTTGTTCCCAATATCAATACCA 183
 Db 609 ACACCCCAACATTTTTCCTTATCTACGACAGCAACATTTTCCCATCCCAACAAACA 668
 OY 184 ACCCCCTACAAACCAACAACGATTC-----CCCAACAACCAACAACGCTTT 234
 Db 669 AACCCCTACAAACCAACAACGATTCGCCCTGACCCCAACCAACAACAACGCTTT 728
 OY 235 CCTGGCCCCCAACAACGATTTCCCTGGCAACCAACAACGATTTCCGACCCCAACAA 294
 Db 729 CCTGAGCCCCCAACAACGATTTCCCGAGAACCAACAACGATTTTCCCGGCAACCA 788
 OY 295 CCAAT-----TCCCAACACACCATTCCTCCCTGGCAACCAACAACGATTTCC 342
 Db 789 CAAATAGTAGCCCAACGACACCAACCAACCATTTCCCTGCAACCAACAACGATTTCT 848
 OY 343 CAGCCCCCAAGAACCAAT-----TCAACAATTAATTTCCGAGCAACCCCAACATCA 393
 Db 849 CAGCCCCCAAGAACCAATTTCTTGGCAACCAACAACCAACCATTTTCGACGCCCTACAACTA 908
 OY 394 TACCCGTGCAACCTCAACAGCAGATTTCTCAACAGCCTCAACCA 438
 Db 909 TAGCCCTTGCAAGCAGCAACACACATTCCTCCCTGGCAACCTCAACTA 953

```

RESULT 9
AB059812      1858 bp      DNA      linear      PLN 03-AUG-2002
LOCUS
DEFINITION    Trilicium aestivum pseudogene for omega gliadin.
ACCESSION     AB059812
VERSION       AB059812.1  GI:16904641
KEYWORDS
SOURCE
ORGANISM      Trilicium aestivum (cultivar:Chinese Spring) DNA.
              Eukaryota; Viridiplantae; Streptophyta; Empidophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Pooideae; Triliceae; Trilicum.
REFERENCE
AUTHORS       Masoudi-Nejad,A., Nasuda,S., Kawabe,A. and Endo,T.R.
TITLE         Molecular cloning, sequencing, and chromosome mapping of a
              1A-encoded Omega-type prolamin sequence from wheat
              Genome 45, 661-669 (2002)
              2 (bases 1 to 1858)
              Masoudi-Nejad,A. and Nasuda,S.
              Direct Submission
              Submitted (12-Apr-2001) All Masoudi-Nejad, Kyoto University,
              Graduate School of Agriculture, Laboratory of Plant Genetics;
              Kitshirakawa olive-cho, Kyoto, Kyoto 606-8502, Japan
              (E-mail:masaoudin@hotmail.com, Tel:81-75-753-6145,
              Fax:81-75-753-6486)
FEATURES
Source
Location/Qualifiers
1..1858
/organism="Trilicium aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/chromosome="1"
/map="1AS"

```


[illegible][illegible]

[illegible][illegible]

FT /transl_except= (pos: 355..357, aa: Glu)
 FT /transl_except= (pos: 358..360, aa: Pro)
 PN WO200002914-A1.
 XX 20-JAN-2000.
 PD 12-JUL-1999; 99WO-AU00563.
 PF 12-JUL-1999; 99WO-AU00563.
 PR 10-JUL-1998; 98AU-0004604.
 XX (CSTR) COMMONWEALTH SCI & IND RES ORG.
 PA (GOOD-) GOODMAN FIELDER LTD.
 PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
 PI Appels R, Morell M, Bekes F, Tamas L;
 DR WPI: 2000-147597/13.
 DR P-PSDB; AAY54568.
 PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
 PT films, coatings, packing materials, adhesives and building materials -
 XX Disclosure; Fig 3; 76pp; English.
 XX
 XX The present sequence encodes a synthetically truncated barley C
 CC hordein protein, which is a glutenin. The protein is designated
 CC ANG-deltacy7cy9236, and is modified, according to the method of
 CC the invention. The specification describes a method for producing a
 CC modified glutenin or seed storage protein, by adding to the protein a
 CC domain that confers the ability to incorporate into gluten, or to bind
 CC a ligand or other macromolecule. The domain can be any domain that will
 CC bind ligands that may be useful in food preparation or in food
 CC composition, e.g. a domain that binds lipids or starch. The method is
 CC used for producing modified glutenins. Glutenins and seed-storage protein
 CC are useful for preparing food products such as leavened or unleavened
 CC breads, pasta, noodles, breakfast cereals, snack foods, cakes, pastries
 CC and foods containing flour based sauces. Glutenins and seed-storage
 CC protein are also useful for preparing nonfood products such as films,
 CC coatings, adhesives, building materials and packaging materials. Grain
 CC or parts of grain containing the modified glutenin and seed-storage
 CC protein is useful for preparing food products. The modified glutenins and
 CC seed-storage proteins are also useful as modifiers of food products in
 CC food industry.
 CC
 SQ Sequence 477 BP; 162 A; 186 C; 43 G; 86 T; 0 other;
 Query Match 100.0%; Score 477; DB 21; Length 477;
 Best Local Similarity 100.0%; Pred. No. 1.4e-100;
 Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 301 CCCCAACAACCATTCCTCCGACACACACACACCATTTCCCAAGCCCAAGAACCAATT 360
 QY 361 CACAATAATATTTTCAGACACACCCCAACATCAACCTGACCACTCAACAGCCATT 420
 DB 361 CACAATAATATTTTCAGACACACCCCAACATCAACCTGACCACTCAACAGCCATT 420
 QY 421 CCTCAACAACCTTCACACAGTCCCAACACAGCTTCATCATATGAGATGCTAG 477
 DB 421 CCTCAACAACCTTCACACAGTCCCAACACAGCTTCATCATATGAGATGCTAG 477
 RESULT 2
 AAV20700
 ID AAV20700 standard; DNA; 5163 BP.
 XX AAV20700;
 AC 17-AUG-1998 (first entry)
 XX
 DT 17-AUG-1998 (first entry)
 XX
 DE Cryptosporidium parvum GP900 antigen open reading frame.
 XX
 KW Open reading frame; ORF; antigen; GP900; Cryptosporidium; infection;
 KW antibody; prophylaxis; treatment; inhibition; retardation;
 KW detection; diagnosis; human; ds.
 XX
 OS Cryptosporidium parvum.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FT 1..5163
 FT CDS
 FT /*tag= a
 FT /*product= GP900 antigen
 FT misc-feature 524..1270
 FT /*tag= b
 FT /*note= "region containing NINC mutations"
 XX
 PN WO9806430-A1.
 XX
 XX 19-FEB-1998.
 PD
 XX 11-AUG-1997; 97WO-US14104.
 PF
 XX 14-AUG-1996; 96US-0700651.
 PR
 XX (RBCC) UNIV CALIFORNIA.
 PA
 XX Gut J, Leech J, Nelson RC, Petersen C;
 PI
 XX
 XX WPI: 1998-159290/14.
 DR P-PSDB; AAM48299.
 DR
 XX
 XX Anti-Cryptosporidium antibody, - used to develop products for
 PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
 PT infections
 PT
 XX
 PS Claim 32; Pages 60-62; 89pp; English.
 XX
 XX The sequence is that encoding the GP900 antigen which may be used
 CC in the production of anti-Cryptosporidium antibodies. These can be
 CC used for the prophylaxis, treatment, inhibition or retardation of
 CC a Cryptosporidium infection in humans or in animals such as calves.
 CC They can also be used for the detection and diagnosis of related
 CC infections.
 CC
 SQ Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 other;
 Query Match 24.9%; Score 118.8; DB 19; Length 5163;
 Best Local Similarity 54.9%; Pred. No. 5.3e-18;
 Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

[illegible]

```
RESULT 3
AAA61849
ID AAA61849 standard; DNA; 5163 BP
```

AC AAA61849;

DT 28-OCT-2000 (first entry)

DE ORF encoding a portion of *Cryptosporidium parvum* NINC isolate GP900

KW GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis; competitive inhibition; attachment; invasion; ligand binding; monoclonal

KW merozoite; diarrhoea; open reading frame; ORF; ds.
 XY

05 *Cryptosporidium parvum*.
XX

EH	Key	Location/Qualifiers
EH	CDS	1669 7182

```

E1      /tag= a
ET      /partial

```

```

ET /product= -Cryptosporidium parvum NINC isolate
ET /note= "No start or stop codons given in the

```

FT specification"

PN US6071518-A.

06-JUN-2000

PF 12-SEP-1997; 97US-0928361.

PR	13-SEP-1996;	96US-0026062.
PR	01-JUN-1993;	93US-0071980

PR	29-MAY-1992;	92US-0891301.
PR	03-APR-1995;	95US-0415751

PR 14-AUG-1996; 96US-0700651.
XY

PA (REGC) UNIV CALIFORNIA.
XX

P1 Petersen C;
XX

DR WFL; 2000-422063/36
DR P-PSDB; AAB11727.

PT New GP900 protein fragments and fusion proteins of *Cryptosporidium*

PT parvum, useful for detecting the presence of the parasite, and
 PT diagnosing or treating Cryptosporidium infections by competitive
 XX inhibition of the function of GP90 -
 Claim 16: Column 47-52; 59pp; English.
 PS

The invention relates to the gp900 glycoprotein of the protozoan *Cryptosporidium parvum*, DNA encoding it, gp900 fragments, and fusion proteins comprising gp900 fragments. The invention also relates to the administration of gp900 or fragments thereof to a host to elicit anti-gp900 antibody production, and to a method of cryptosporidiosis treatment or prophylaxis comprising administration of anti-gp900 antibodies to an individual. *Cryptosporidium parvum* gp900 and gp900 fragments are able to competitively inhibit sporozoite or merozoite attachment or invasion, and are also useful for the generation of anti-gp900 antibodies. The antibodies also inhibit sporozoite or merozoite attachment/invasion and additionally inhibit the binding of gp900 ligands to gp900. gp900 proteins, fragments and antibodies may therefore be used to treat or prevent cryptosporidiosis. Infection with *Cryptosporidium* is a common cause of diarrhoea in humans and causes life-threatening diarrhoea in immunocompromised persons. *Cryptosporidiosis* can be contracted from contaminated municipal water supplies (e.g., public swimming pools). It is also a cause of disease in animals, resulting in financial losses in agriculture. gp900 fragments, fusion proteins and antibodies may also be used for the diagnosis of *Cryptosporidium parvum* infections, and for the detection of the parasite in the environment. The present sequence represents the open reading frame (ORF) encoding a portion of the gp900 protein of the N10c isolate of *Cryptosporidium parvum*.

SQ Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other.

Query Match	24.98;	Score 118.8;	DB 21;	Length 5163;
Query Match	24.98;	Score 118.8;	DB 21;	Length 5163;

Matches	234;	Conservative	0;	Mismatches	192;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	------	--------	----	------	----

[illegible]

RESULTS A

ABT04778 standard: DNA: 5163 BP

AA ABT04778

FF	27-SEP-2002 (first entry)
XX	C parvum GP900 gene fragment SEQ ID NO: 4.
DE	Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
XX	gene; ds.
KW	Cryptosporidium parvum.
RN	MO200194631-A1.
XX	
PD	13-DEC-2001.
XX	
PF	14-MAY-2001; 2001WO-US15624.
XX	
PR	06-JUN-2000; 2000US-0588995.
XX	(REGC) UNIV CALIFORNIA.
PA	Petersen C, Barnes DA, Nelson RG, Gut J;
PI	WPI; 2002-566447/60.
DR	Detecting Cryptosporidium in biological and environmental samples and
XX	diagnosis of cryptosporidiosis involves, contacting the sample with
PT	Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
PS	-
XX	Disclosure: Page 103-104; 157pp; English.
CC	The present invention relates to a method of detecting Cryptosporidium in
CC	biological and environmental samples, and of diagnosing
CC	cryptosporidiosis. This involves obtaining a sample and contacting it
CC	with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC	RNA, or its variant, mutant or fragment. The method is also useful for
CC	detecting and identifying individual Cryptosporidium isolates based on
CC	the genetic characteristics, and for diagnosis of prior or concurrent
CC	Cryptosporidium infection. The present sequence is a C. parvum coding
CC	sequence used in the exemplification of the invention.
XX	
SQ	Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other;
Query Match	24.9%; Score 118.8; DB 24; Length 5163;
Best Local Similarity	54.9%; Pred. No. 5.3e-18;
Matches 234; Conservative	0; Mismatches 192; Indels 0; Gaps 0;
OY	25 CAAGATTGCGCATTCACCAACAATATCTGTGGCGGCATATCCAAANAACCATAT 84
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	680 CAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAGACTA 739
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	85 CTACGCCAAAACCATTTCCAGTGCAGCACCGCTTTCACACACCCCACACAATTTCCCC 144
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 799
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	145 TATCTACGAGGAAGATTGTTTCCCAATTCAAATATACCAACCCTTCAACACAGAACA 204
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	800 CAACTTACACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 859
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	205 CCATTTCGCCAACACCAACACACCTTCTTGCGGCCCAACAACAATTTCCCTGGCNA 264
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	860 CTACAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 919
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	265 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 324
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	920 CACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 979
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	325 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 384
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	980 CAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1033
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY	385 CAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 444
DB	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

[illegible]

```
QY 145 TATCTACGAGGAATGTTTCCCAATATGCAATACCAACCCCTACACCAACAA 204
    |||||
DB 801 CAACTACAAACCAACAACTACAAACCAACCAACCAACCAACCAACCA 860
QY 205 CCATTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 264
    |||||
DB 861 CTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 920
    |||||
QY 265 CCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 324
    |||||
DB 921 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 980
    |||||
QY 325 CCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 384
    |||||
DB 981 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1040
    |||||
QY 385 CAACATCATACCTGTGCAACCTGCAACGCAATTTCTCTCAACCAACCAACCA 444
    |||||
DB 1041 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1100
    |||||
QY 445 CAACAA 450
    |||||
DB 1101 CAACAA 1106
    |||||

RESULT 6
AAA61848
ID AAA61848 standard; DNA; 5318 BP.
AC AAA61848;
XX
DT 28-OCT-2000 (first entry)
XX
DE DNA encoding a portion of Cryptosporidium parvum NINC isolate GP900.
XX
KW GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;
KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW merozoite; diarrhoea; protozoacide; ds.
XX
OS Cryptosporidium parvum.
XX
XX Key Location/Qualifiers
FH 1669..7182
FT CDS
FT /*tag= a
FT /partial
FT /product= "Cryptosporidium parvum NINC isolate GP900"
FT /note= "No start codon given in the specification"
FT 5167..5318
FT /*tag= b
XX
XX 3'UTR
XX
XX US6071518-A.
XX
XX 06-JUN-2000.
XX
XX 12-SEP-1997; 97US-0928361.
XX
XX 13-SEP-1996; 96US-0026062.
XX 01-JUN-1993; 93US-0071880.
XX 29-MAY-1992; 92US-0891301.
XX 03-APR-1995; 95US-0415751.
XX 14-AUG-1996; 96US-0700651.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Petersen C;
XX
XX WPI: 2000-422065/36.
XX P-PSDB; AAB11727.
XX
XX New GP900 protein fragments and fusion proteins of Cryptosporidium
XX parvum, useful for detecting the presence of the parasite, and
XX diagnosing or treating Cryptosporidium infections by competitive
XX inhibition of the function of GP900 -
```

```
XX
PS Claim 16; Column 41-48; 59pp; English.
XX
CC The invention relates to the GP900 glycoprotein of the protozoan
CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC proteins comprising GP900 fragments. The invention also relates to the
CC administration of GP900 or fragments thereof to a host to elicit anti-
CC GP900 antibody production, and to a method of cryptosporidiosis treatment
CC or prophylaxis comprising administration of anti-GP900 antibodies to an
CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC competitively inhibit sporozoite or merozoite attachment or invasion, and
CC are also useful for the generation of anti-GP900 antibodies. The
CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC additionally inhibit the binding of GP900 ligands to GP900.
CC proteins, fragments and antibodies may therefore be used to treat or
CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC immunocompromised persons. Cryptosporidiosis can be contracted from
CC contaminated municipal water supplies (e.g., public swimming pools). It
CC is also a cause of disease in animals, resulting in financial losses in
CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC used for the diagnosis of cryptosporidium parvum infections, and for the
CC detection of the parasite in the environment. The present sequence
CC represents genomic DNA encoding a portion of the GP900 protein of the
CC NINC isolate of Cryptosporidium parvum.
XX
XX Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 other;
XX
XX
XX Query Match 24.9%; Score 118.8; DB 21; Length 5318;
XX Best Local Similarity 54.9%; Pred. No. 5.4e-18;
XX Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
XX
QY 25 CAAGAGTTGCAATGCAACCAACCAATATCATGCGCCGCAACCAATCCCAAAACCAATAT 84
    |||||
DB 680 CAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 739
    |||||
QY 85 CTACCGCAAAACCAATTTCCAGTGCAACCGTTTTCACACACCCCAACCAATATTTCC 144
    |||||
DB 740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 799
    |||||
QY 145 TATCTACGAGGAATGTTTCCCAATATGCAATACCAACCCCTACACCAACAA 204
    |||||
DB 800 CAACTACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 859
    |||||
QY 205 CCATTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 264
    |||||
DB 860 CTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 919
    |||||
QY 265 CCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 324
    |||||
DB 920 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 979
    |||||
QY 325 CCACACCAACCAATTTCCCAACCAACCAACCAACCAATTAATTTCCAGCAACCC 384
    |||||
DB 980 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1039
    |||||
QY 385 CAACATCATACCTGTGCAACCTGCAACGCAATTTCTCTCAACCAACCAACCA 444
    |||||
DB 1040 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1099
    |||||
QY 445 CAACAA 450
    |||||
DB 1100 CAACAA 1105
    |||||

RESULT 7
ABT04777
ID ABT04777 standard; DNA; 5318 BP.
AC ABT04777;
XX
XX 27-SEP-2002 (first entry)
XX
```

DE C parvum GP900 gene fragment SEQ ID NO: 3.
 XX Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
 KW gene; ds.
 XX Cryptosporidium parvum.
 OS
 PN WO200194631-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 14-MAY-2001; 2001WO-US15624.
 XX
 PR 06-JUN-2000; 2000US-0588995.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Petersen C, Barnes DA, Nelson RG, Gut J;
 XX
 DR WPI: 2002-566447/60.
 XX
 XX Detecting Cryptosporidium in biological and environmental samples and
 PT diagnosis of cryptosporidiosis involves, contacting the sample with
 PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
 PT
 XX
 PS Disclosure: Page 101-103; 157pp; English.
 XX
 XX The present invention relates to a method of detecting Cryptosporidium in
 CC biological and environmental samples, and of diagnosing
 CC cryptosporidiosis. This involves obtaining a sample and contacting it
 CC with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
 CC RNA, or its variant, mutant or fragment. The method is also useful for
 CC detecting and identifying individual Cryptosporidium isolates based on
 CC the genetic characteristics, and for diagnosis of prior or concurrent
 CC cryptosporidium infection. The present sequence is a C. parvum coding
 CC sequence used in the exemplification of the invention.
 XX
 SQ Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 other:
 Query Match 24.9%; Score 118.8; DB 24; Length 5318;
 Best Local Similarity 54.9%; Pred. No. 5.4e-18;
 Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

DB 1100 CACAA 1105
 |||||
 RESULT 8
 AAA61847
 ID AAA61847 standard; DNA; 5511 BP.
 XX
 AC AAA61847;
 XX
 DT 28-OCT-2000 (first entry)
 XX
 DE Cryptosporidium parvum Iowa isolate GP900 ORF.
 XX
 KW GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
 KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
 KW merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.
 XX
 OS Cryptosporidium parvum.
 XX
 XX Key Location/Qualifiers
 FT CDS 1.5511
 FT /tag= a
 FT /product= "Cryptosporidium parvum Iowa isolate GP900"
 FT /note= "No stop codon given"
 XX
 PN US6071518-A.
 XX
 PD 06-JUN-2000.
 XX
 PF 12-SEP-1997; 97US-0928361.
 XX
 PR 13-SEP-1996; 96US-0026062.
 PR 01-JUN-1993; 93US-0071880.
 PR 29-MAY-1992; 92US-0891301.
 PR 03-APR-1995; 95US-0415751.
 PR 14-AUG-1996; 96US-0700651.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 PI Petersen C;
 XX
 DR WPI: 2000-422065/26.
 DR P-PSDB; AAb11726.
 XX
 XX New GP900 protein fragments and fusion proteins of Cryptosporidium
 PT parvum, useful for detecting the presence of the parasite, and
 PT diagnosing or treating Cryptosporidium infections by competitive
 PT inhibition of the function of GP900 -
 XX
 PS Claim 16; Column 37-42; 59pp; English.
 XX
 CC The invention relates to the GP900 glycoprotein of the protozoan
 CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
 CC proteins comprising GP900 fragments. The invention also relates to the
 CC administration of GP900 or fragments thereof to a host to elicit anti-
 CC GP900 antibody production, and to a method of cryptosporidiosis treatment
 CC or prophylaxis comprising administration of anti-GP900 antibodies to an
 CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
 CC competitively inhibit sporozoite or merozoite attachment or invasion, and
 CC are also useful for the generation of anti-GP900 antibodies. The
 CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
 CC additionally inhibit the binding of GP900 ligands to GP900.
 CC proteins, fragments and antibodies may therefore be used to treat or
 CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
 CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
 CC immunocompromised persons. Cryptosporidiosis can be contracted from
 CC contaminated municipal water supplies (e.g., public swimming pools). It
 CC is also a cause of disease in animals, resulting in financial losses in
 CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
 CC used for the diagnosis of Cryptosporidium parvum infections, and for the
 CC detection of the parasite in the environment. The present sequence
 CC represents the open erading frame (ORF) encoding the GP900 protein of the

QY	214	CAACGACCAACGACACCTTCTGCGGCCCAACAAACATTCCTCGGACACCAACAA	273
Db	2894	CTACTACGACACACACCAACACCAACCACTACCAAGAAACCAACCAACCAACAA	295
QY	274	CCATTTCGCCAGCCCAAGAACCAATTCGCCAACCAACATTCCTCGGACACCAACAA	333
Db	2954	CACACAAACAAACAACTACTACTACACACGACAAACAAACCAACGACACCAACAA	301
QY	334	CCATTTCGCCAGCCCAAGAACCAATTCACCAATTAATTTTCAGCAACCCCAACATCA	393
Db	3014	CAACTACGACACCAAGAAACCAACCACTACTACTACTACTACCAACCAACCAACAACTA	307
QY	394	TACCTGTGCAACCTCAACAGCCATTTCTCTCAACCAACCTCAACGAGTCCGCCAACAA	450
Db	3074	CTACTACTACGACACCAACCAACCACTACTACTACTACTACTACCAACCAACCAACCA	3130
RESULT 12			
AAV38816			
ID	AAV38816	standard; DNA; 1107 BP.	
AC	AAV38816;		
XX	09-OCT-1998	(first entry)	
DT			
XX	DNA encoding a durum wheat glutenin protein.		
DE			
XX	Glutenin gene: durum wheat; low-molecular-weight;		
KW	transgenic durum wheat; ss.		
KW			
XX	Triticum durum.		
OS			
XX	FR2757538-A1.		
PN			
PD	26-JUN-1998.		
XX			
PP	18-DEC-1997; 97FR-0016059.		
XX			
PR	19-DEC-1996; 96IT-MI02663.		
XX			
PA	(ITU-Y-) ITAL MIN UNIV RICECA SCI & TECNOLOGICA.		
XX			
PI	Cardelli LE, D'Ovidio R, Marchitelli C, Porceddu E;		
XX			
DR	WPI; 1998-365055/32.		
DR	P-PSDB; AAM62647.		
XX			
PT	Durum wheat glutenin gene - coding for glutenin protein of low		
PT	molecular weight		
XX			
PS	Claim 1; Page 13; 18pp; French.		
XX			
CC	The present sequence represents a glutenin gene, and is isolated from		
CC	the genomic DNA of Triticum durum L. The gene codes for a		
CC	low-molecular-weight glutenin protein and can be used to produce		
CC	transgenic durum wheat plants with "better quality characteristics"		
CC	[no details given].		
XX			
SQ	Sequence 1107 BP: 368 A; 377 C; 152 G; 210 T; 0 other;		
Query Match 22.5%; Score 107.4; DB 19; Length 1107;			
Best Local Similarity 56.7%; Pred. No. 1.6e-15;			
Matches 242; Conservative 0; Mismatches 176; Indels 9; Gaps 2;			
QY	33	GCAATCACACACATCATATCTGCGCGACGACCATATCCACAAACCATATCTACCGCA	92
Db	135	GCAACAAACAAACACACCATATATGCAACAAACAAACACCATTTTGGACGAACAA	194
QY	93	AAAACCATTTCCAGTGTG---CAGCAACGCTTTCACACACCCCAACAAATATTTCCCTATCT	149
Db	195	ACCACCATTTTTCACGAGCAACCAACCAAGTGTTCACCGCAACCAACATATTTTGGACAA	254
QY	150	ACGAGAGAAATGTTTCCCAATATCAAAATACCAACCCCTTACACGACAAACCAATT	209

Db	255	ACAACTACGACCAATTTTCGAGACACACACACACACACCACTG	314
OY	210	CCCCCAACACCAACACACCTTCTTCGCGCCCAACACCAATTCCTGCAACCA	269
Db	315	TCTACCGCAACACCACTATTTTCGAGACACACACACACACCACTTTTCAGCAACTAC	374
OY	270	ACAACCAATTTCCCGACGCCCAAGAACCAATTTCCCAACACCAATTCCTGGCAACCA	329
Db	375	ACCAATTTTCGACGACACACACAGTACTACGCAACACACCAATTTTCGACGACCAAC	434
OY	330	ACAACCAATTTCCCGACGCCCAAGAACCAATTTTCGACGACCAAC	383
Db	435	ACCAACCAATTTTCGACGACCACTACCACTTTTCGAGACACACACACCACTGATACCGCA	494
OY	384	CAACCAATTTACCTGCGACCACTGACAGCAATTTCTTCAACACCTCAACCACTGCC	443
Db	495	ACAACCACTTTTCGACGACCACTGACCACTGACCACTGACCACTGACCACTTTTC	554
OY	444	CCAACCA 450	
Db	555	GCAACCA 561	
RESULT 13			
ID	ABA49946		
XX	ABA49946	standard; DNA: 1635 BP.	
AC	ABA49946;		
XX	DT	01-FEB-2002 (first entry)	
DE		Human breast cell single exon nucleic acid probe #8641.	
XX		Human; microarray: single exon probe; gene expression: breast;	
KW		disease; cancer; ss.	
XX		Homo sapiens.	
OS			
XX		MO200157271-A2.	
PN		09-AUG-2001.	
PD			
XX			
PE		30-JAN-2001; 2001MO-US00662.	
XX			
PR		04-FEB-2000; 2000US-0180312.	
PR		26-MAY-2000; 2000US-0207456.	
PR		30-JUN-2000; 2000US-0608408.	
PR		03-AUG-2000; 2000US-0632366.	
PR		21-SEP-2000; 2000US-0234687.	
PR		27-SEP-2000; 2000US-0236359.	
PR		04-OCT-2000; 2000GB-0024263.	
XX			
PA		(MOLE-) MOLECULAR DYNAMICS INC.	
XX			
PI		Penn SG, Hanzel DK, Chen W, Rank DR;	
XX			
DR		WPI: 2001-496933/54.	
XX			
PT		New spatially-addressable set of single exon nucleic acid probes,	
PT		useful for measuring gene expression in sample derived from human	
PT		breast, comprises number of single exon nucleic acid probes -	
XX			
PS		Claim 4; SEQ ID NO 8641; 327bp + sequence listing; English.	
CC		The invention relates to a spatially-addressable set of single exon	
CC		nucleic acid probes for measuring gene expression in a sample derived	
CC		from human breast and BT 474 cells. The method involves contacting	
CC		the probes with a collection of detectably labelled nucleic acids	
CC		derived from mRNA of human breast, and then measuring the label	
CC		bound to each probe of the microarray. The probes are useful for	
CC		verifying the expression of regions of genomic DNA predicted to	
CC		encode proteins. They are useful for gene discovery, and for	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:55:25 ; Search time 51.7974 Seconds
(without alignments)
2824.173 Million cell updates/sec

Title: US-09-743-533-18
Perfect score: 477
Sequence: 1 atgagcgaactaaacctg.....gcatalgagatgctctag 477

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCtUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfilltest.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118.8	24.9	5163	3 US-08-700-651-1	Sequence 1, Appl
2	118.8	24.9	5163	3 US-08-928-361B-4	Sequence 4, Appl
3	118.8	24.9	5318	3 US-08-700-651-2	Sequence 2, Appl
4	118.8	24.9	5318	3 US-08-928-361B-3	Sequence 3, Appl
5	111.4	23.4	5511	3 US-08-928-361B-2	Sequence 2, Appl
6	111.4	23.4	7334	3 US-08-928-361B-1	Sequence 1, Appl
7	107.4	22.5	1107	2 US-08-991-300-1	Sequence 1, Appl
8	73.2	15.3	2455	4 US-09-103-429A-1	Sequence 1, Appl
9	73.2	15.3	2821	4 US-09-103-429A-2	Sequence 2, Appl
10	71	14.9	8442	4 US-09-272-032-6	Sequence 6, Appl
11	70.6	14.8	688	4 US-08-998-416-915	Sequence 915, App
12	70.6	14.8	10747	2 US-08-452-872-1	Sequence 1, Appl
13	70.6	14.8	10747	5 US-08-452-872-1	Sequence 1, Appl
14	70.6	14.8	10747	5 PCT-US93-03985-1	Sequence 1, Appl
15	70.6	14.8	24979	2 US-08-147-777-3	Sequence 1, Appl
16	70.6	14.8	24979	3 US-08-452-872-3	Sequence 3, Appl
17	70.6	14.8	24979	5 PCT-US93-03985-3	Sequence 3, Appl
18	68.8	14.4	198	5 PCT-US95-10668-3	Sequence 3, Appl
19	68.8	14.4	198	5 PCT-US95-10668-4	Sequence 4, Appl
20	67.2	14.1	198	5 PCT-US95-10668-1	Sequence 1, Appl
21	67.2	14.1	198	5 PCT-US95-10668-2	Sequence 2, Appl
22	66.8	14.0	1505	1 US-07-915-246-1	Sequence 1, Appl
23	60	12.6	925	3 US-08-858-003-1	Sequence 1, Appl
24	60	12.6	925	3 US-09-078-166-1	Sequence 1, Appl
25	60	12.6	925	4 US-08-997-467-1	Sequence 1, Appl
26	57.8	12.1	1086	1 US-08-415-751-47	Sequence 47, Appl
27	57.8	12.1	1608	4 US-09-292-225-20	Sequence 20, Appl

C	28	57.8	12.1	1608	4 US-09-292-225-22	Sequence 22, Appl
	29	57.8	12.1	1665	4 US-09-292-225-17	Sequence 17, Appl
C	30	57.8	12.1	1665	4 US-09-292-225-19	Sequence 19, Appl
	31	57.8	12.1	1752	4 US-09-292-225-14	Sequence 14, Appl
C	32	57.8	12.1	1752	4 US-09-292-225-16	Sequence 16, Appl
	33	57	11.9	1690	1 US-08-276-452A-24	Sequence 24, Appl
	34	57	11.9	1690	2 US-08-798-744-24	Sequence 24, Appl
	35	57	11.9	2032	4 US-09-241-581B-5	Sequence 5, Appl
	36	57	11.9	2032	5 US-08-265-428-5	Sequence 5, Appl
	37	57	11.9	2032	5 PCT-US95-07721-5	Sequence 5, Appl
	38	56.6	11.9	7785	2 US-08-276-967-1	Sequence 1, Appl
	39	53.8	11.3	2824	4 US-07-757-022B-13	Sequence 13, Appl
	40	53.8	11.3	3066	4 US-07-757-022B-83	Sequence 83, Appl
	41	53.8	11.3	3117	4 US-07-757-022B-73	Sequence 73, Appl
	42	53.8	11.3	3148	4 US-07-757-022B-57	Sequence 57, Appl
	43	53.8	11.3	3420	4 US-07-757-022B-103	Sequence 103, Appl
	44	53.8	11.3	3813	4 US-07-757-022B-43	Sequence 43, Appl
	45	53.8	11.3	3936	4 US-07-757-022B-41	Sequence 41, Appl

ALIGNMENTS

```
RESULT 1
US-08-700-651-1
: Sequence 1, Application US/08700651B
: Patent No. 6015882
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUY, JIRI
: TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
: TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
: FILE REFERENCE: 480.19-4(HV)
: CURRENT APPLICATION NUMBER: US/08/700,651B
: CURRENT FILING DATE: 1997-08-14
: EARLIER APPLICATION NUMBER: 08/415,751
: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 5163
: TYPE: DNA
: ORGANISM: Cryptosporidium parvum
: US-08-700-651-1
:
Query Match      24.9%; Score 118.8; DB 3; Length 5163;
Best Local Similarity 54.9%; Pred. No. 2.2e-20;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

25 CAAGAGTTCATATCCACCAACATATATCTGCCGACCATATTCACCAAAACCATAT 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
681 CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85 CTACCCGCAAAACATATTCACAGCGACCAACCTTTCACACACCCCAACAATATTTCC 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
741 CTACTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 TATCTACGAGAGATTTGTTCCCAATATCAATATCAACACCCCTACACACACAA 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 CAACCTACACACACACACACACACACACACACACACACACACACACACACACAC 860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 CCATTTCACACACACACACACACACCTTTCTCGGCCCCCAACACACATTTCCCTGG 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
861 CTACACACACACACACACACACACACACACACACACACACACACACACACACAC 920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 CCACACACACATTTCCCGCCGACCAACCAATTTCCCAACACACATTTCCCTGGCA 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
921 CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325 CCACACACACATTTCCCGCCGACCAACCAATTTCAACCAATATTTCCAGCACCC 384
```

Db 981 CAACAAACAAACAACTACACTACAGAAACCAACACTACTACTACTACAA 1040
OY 385 CAACAATCATCTGCTGGCAACCTCAACAGCCATTTCTTCAACACTCAACGATGCC 444
Db 1041 CAACAACAACTACTACTACTACTACCAACAACTACTACTACTACTACAA 1100
OY 445 CAACAA 450
Db 1101 CAACAA 1106

RESULT 2
US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480,76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5163 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-928-361B-4

Query Match 24.9%; Score 118.8; DB 3; Length 5163;
Best Local Similarity 54.9%; Pred. No. 2.2e-20;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

OY 25 CAAGAGTGCATATCCACACATATATATGCGCGAGCCATATCCACAAACCCATAT 84
Db 680 CAACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 739
OY 85 CTACCGCAAAACCAATTTCCAGTGCAGACCGTTTACACACACCCCAACATATTTCCC 144
Db 740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACA 799
OY 145 TATCTACAGAGAAATTTGTTCCCAATATATACCAACCCCTTACCAACCAACA 204
Db 800 CAACTACAAACCAACCACTACACACCAACCAACCAACCAACCAACCAACCA 859

OY 205 CCATTGCCCAACCAACCAACCAACCTCTGCTGGCCCAACCAACCAATTCCTGGCAA 264
Db 860 CTACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 919
OY 265 CCACAACCAACCAATTTCCAGGCCCCAAGAACCAATTTCCCAACCAACCAATTTCCG 324
Db 920 CAACAACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 979
OY 325 CCACAACCAACCAATTTCCAGGCCCCAAGAACCAATTTCAACCAATTTTTCAGCAACC 384
Db 980 CAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1039
OY 385 CAACAATCATCTGCTGGCAACCTCAACAGCCATTTCTTCAACCACTCAACGATGCC 444
Db 1040 CAACAACCAACCACTACTACTACTACTACCAACAACTACTACTACTACTACA 1099
OY 445 CAACAA 450
Db 1100 CAACAA 1105

RESULT 3
US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480,19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-2

Query Match 24.9%; Score 118.8; DB 3; Length 5318;
Best Local Similarity 54.9%; Pred. No. 2.2e-20;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

OY 25 CAAGAGTGCATATCCACACATATATGCGCGAGCCATATCCACAAACCCATAT 84
Db 681 CAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 740
OY 85 CTACCGCAAAACCAATTTCCAGTGCAGACCGTTTACACACACCCCAACATATTTCCC 144
Db 741 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACA 800
OY 145 TATCTACAGAGAAATTTGTTCCCAATATATACCAACCCCTTACCAACCAACA 204
Db 801 CAACTACAAACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCA 860
OY 205 CCATTGCCCAACCAACCAACCAACCTCTGCTGGCCCAACCAACCAATTCCTGGCAA 264
Db 861 CTACAAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACA 920
OY 265 CCACAACCAACCAATTTCCAGGCCCCAAGAACCAATTTCCCAACCAACCAATTTCCG 324
Db 921 CAACAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACA 980
OY 325 CCACAACCAACCAATTTCCAGGCCCCAAGAACCAATTTCAACCAATTTTTCAGCAACC 384

Db 981 CAACAACACACACACACTAGCACTACCAAGAACCACTACTACTACTACCA 1040
Oy 385 CAACAATCATACCTGTGCAACCTCAACAGCAATTTCTTCAACAACCTCAACGAGTCCC 444
Db 1041 CAACAACACACACTACTACTACTACTACCAACACACACTACTACTACTACTACCA 1100
Oy 445 CAACAA 450
Db 1101 CAACAA 1106

RESULT 4

US-08-928-361B-3
Sequence 3, Application US/08928361B
Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:

NAME: VERNY, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5318 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3

Query Match 24.9% Score 118.8; DB 3; Length 5318;
Best Local Similarity 54.9% Pred. No. 2.2e-20;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

Oy 25 CAAGAGTTGCATTCACACACATATCTGCGGAGCCATATCCACAAACCCATAT 84
Db 680 CAACAAC 739
Oy 85 CTACCGCAAAACCATTTCCAGTCCAGCACCGTTTCACACACCCCAACATATTTCCC 144
Db 740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 799
Oy 145 TATCTACACAGAGATTTGTTCCCAATATCAATATACACACCCCAACACACAC 204
Db 800 CAATAC 859

Oy 205 CCATTTCCCCAACACACACACACACTCTTCTCTGGCCCCCAACACACATTCCTGGCAA 264
Db 860 CTACACACACACACACACTACACACACACACACACACACACACACACACAC 919
Oy 265 CCACACACACACATTTCCCGCCGCAAGACCAATTTCCCAACACACATTCCTGGCAA 324
Db 920 CAACAACACACACACTACACACACACACACACACACACACACACTACTACTAC 979
Oy 325 CCACACACACACATTTCCCGCCGCAAGACCAATTTCAACATATATTTCCAGCACCC 384
Db 980 CAACAACACACACACACACTACTACTACTACTACTACTACTACTACTACTACT 1039
Oy 385 CAACAATCATACCTGTGCAACCTCAACAGCAATTTCTTCAACAACCTCAACGAGTCCC 444
Db 1040 CAACAACACACACTACTACTACTACTACTACCAACACACACACTACTACTACTAC 1099
Oy 445 CAACAA 450
Db 1100 CAACAA 1105

RESULT 5

US-08-928-361B-2
Sequence 2, Application US/08928361B
Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:

NAME: VERNY, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-928-361B-2

Query Match 23.4% Score 111.4; DB 3; Length 5511;
Best Local Similarity 54.2% Pred. No. 1.5e-18;
Matches 226; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

Oy 34 CAATACACACACACATATATCTGCGGAGCCATATTCACAAACCCATATATCTACCGCA 93
Db 111 11111111 11 11 11 11111 11 11 11

```

Db      1046 CAACACACACAACTACTACTACTACTAGAACCCACAACTACAAACCAAGCAACTA 1105
OY      94 AAACCATTTCCAGTCAGACACCGTTTCACACACCCCAACATATTTCCCTATCTACCA 153
Db      1106 CAACCAACACACCACTACTACTACTACTAGAAACCAACACACACAAACAACTACTACTA 1165
OY      154 GAGGAATTTGTTCCCAATATTCAAATACCAACCCCTTACACACCAACAAACAACTTCCCC 213
Db      1166 CTACAAACAAACACACACACACTACTACTACTACTACTACTACTACTACTACTACTACTA 1225
OY      214 CAACACACACAAACAACTCTTCTCTGCGCCCAACACCACTTCCCTGGCAACCAACAA 273
Db      1226 CTACTACACAAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 1285
OY      274 CCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCACTTCCCTGGCAACCAACAA 333
Db      1286 CAACACACACAAACAACTACTACTACTACTAGACACACACAAACCAACCAACCAACCA 1345
OY      334 CCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCACTTCCCTGGCAACCAACAA 393
Db      1346 CAACACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1405
OY      394 TACCCGTGCAACCTCAACGAGCATTTCTCAACCACTCAACCAAGTCCCAACAA 450
Db      1406 CTACTACTACCAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTA 1462

```

RESULT 6

```

US-08-928-361B-1
; Sequence 1, Application US/08928361B
; Patent No. 6071518

```

GENERAL INFORMATION:

```

; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480,76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 734 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-928-361B-1

```

```

Query Match      23.4%; Score 111.4; DB 3; Length 734;
Best Local Similarity 54.2%; Pred. No. 1,6e-18;
Matches 226; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

OY      34 CAATCACCACACAACTATCTGCGGACCATATTCACAAACCCATATCTACCGCA 93
Db      2714 CAACACACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 2773
OY      94 AAACCATTTCCAGTCAGACACCGTTTCACACACCCCAACATATTTCCCTATCTACCA 153
Db      2774 CAACCAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 2833
OY      154 GAGGAATTTGTTCCCAATATTCAAATACCAACCCCTTACACACCAACCAACCACTTCCCC 213
Db      2834 CTACAAACAAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 2893
OY      214 CAACACACACAAACAACTCTTCTCTGCGCCCAACACCACTTCCCTGGCAACCAACAA 273
Db      2894 CTACTACACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 2953
OY      274 CCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCACTTCCCTGGCAACCAACAA 333
Db      2954 CAACACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 3013
OY      334 CCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCACTTCCCTGGCAACCAACAA 393
Db      3014 CAACCTACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 3073
OY      394 TACCCGTGCAACCTCAACGAGCATTTCTCAACCACTCAACCAAGTCCCAACAA 450
Db      3074 CTACTACTACCAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTA 3130

```

RESULT 7

```

US-08-991-300-1
; Sequence 1, Application US/08991300
; Patent No. 5973225

```

GENERAL INFORMATION:

```

; APPLICANT: D'OVIDIO, RENATO
; APPLICANT: PORCEDDU, ENRICO
; APPLICANT: MERCHITELLI, CINZIA
; APPLICANT: CARDELLI, LUISA ERCOLI
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
; TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,300
; FILING DATE: 16-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IT MI 96/A 002663
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2264-0201-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-991-300-1

```

Query Match	22.5%;	Score 107.4;	DB 2;	Length 1107;
Best Local Similarity	56.7%;	Pred. No. 8.6e-18;		
Matches: 242;	Conservative	0;	Mismatches 176;	Indels 9;
				Gaps 2

QY	33	GCATATCCCAACAACATCATATCTGCGCAGCATATCCAAAAACCATATTTACCGCA	92
Db	135	GCACAACAACAACCAACCATTTATGGCAACACACACACACACCATTTTTCGACGACAACA	194
QY	93	AAAACATTTTCCAGTg---CGAACACGGTTTCACACACCCGACAGCAATATTTCCCTATCT	149
Db	195	ACCACCATTTTTCACACACACACACAGTTTCTACCGCAACACCAATCATTTTTCGACGCA	254
QY	150	ACCAGAGGAATGTGTTTCCCATATTCAAATACCAACCCCTTACACACCAACACCATTT	208
Db	255	ACAAGTACCAACATTTTTCGACACAAACAACACCAATTTTTCACACGACAACAACCAAGT	314
QY	210	CCCCAACAACAACAACACTCTTCTCTGGGCCCAACAACATTTCCCTGGCAACACA	268
Db	315	TTTACGGCAACAACATCATTTTTCGACAGAACATACACACATTTTTCACACGCAACTAC	374
QY	270	ACAACCATTTTCCCGACGCCCAAGAAACAATTTCCCAACACACATTTCCCTGGCAACACA	329
Db	375	ACCATTTTTCGACGACAACAACAGTACTACCGCAACACACCATTTTTCGACGACAAC	434
QY	330	ACAAACATTTTCCCGACGCCCAAGAAACAATTTT-----CAACAATTAATTTTTCGCAAC	383
Db	435	ACACACATTTTTCACACACACACTACCAACATTTTTCGACGACAACAACAACAGTACTACCGCA	494
QY	384	CCAACATCATACCTGTGTGACACTCAACAGGCATTTTCTCAACACCTCAACAGTCTCC	443
Db	495	ACAACCAACATTTTTCGACACACACACACAAACAAATTTTCACCGCAACACACACATTTTTC	554
QY	444	CCAACA 450	
Db	555	GCACAA 561	

RESULT 8
US-09-103-429A-1
Sequence 1, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin1
TITLE OF INVENTION: CNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tlaga
City: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A

```

1 REGISTRATION NUMBER: 34.390
2 REFERENCE/POCKET NUMBER: BTL-39
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (607) 256-2000
5 TELEFAX: (607) 256-3628
6 INFORMATION FOR SEQ ID NO: 1:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 2455 base pairs
9 TYPE: nucleic acid
10 STRANDEDNESS: double
11 TOPOLOGY: linear
12 MOLECULE TYPE: cDNA
13 HYPOTHEetical: NO
14 ANTI-SENSE: NO
15 FRAGMENT TYPE: N-terminal
16 ORIGINAL SOURCE:
17 ORGANISM: Trichopustia n1
18 TISSUE TYPE: peritrophic Membrane
19 IMMEDIATE SOURCE:
20 CLONE: IM14
21 JS-09-103-429A-1

```

Query Match	15.38;	Score 73.2;	DB 4;	length 2455;
Best Local Similarity	49.08;	Pred. No. 2.7e-09;		
Matches 195; Conservative	0;	Mismatches 203;	Indels 0;	Gaps 0

OY		5	CTGCGCCAGCAGATATTCACAAAACCATTATACGGCAAAAAACATTTTCCAGTGACAA	114
Db		291	CGCCTTTTAGCGGATGTACCTCTGGCCAAGACCTCCAGCTGAGACAACCCAGGCCCAAGA	350
OY		115	CCGTTTTACAGACCCCAACCATATTTTCCCCTATTACGACGAGAGGAATTTTCCCCAATAT	174
Db		351	CAACTCAGGCGCCCAACCAACCAACCCAGGCGCCCAACGACAATCTAGGCGCCCTACTACAA	410
OY		175	CAATTTACCAACCCCTCTACACCAACACCAATTTCCCCCAACACCAACAACACTCTT	234
Db		411	CCACCCAGGCGCCCAACCAACCAACCCAGGCGCCCAACCAACCAACCAACCAACCA	470
OY		235	CGATGGCCCAACCAACCATTTTCCGCGCAACCAACAACCATTTTCCCGACGCCCAAGA	294
Db		471	CCGAGGCGCCCAACTTACCACTCAGGCCCTTACTACTACCACTGAGGCCCAACCAACA	530
OY		295	CCAAATTTCCCAACCAACCATTTTCCCTGGCAACCAACAACCATTTTCCCGACGCCCAAGA	354
Db		531	CTGAGGCGCCCTACACCAACCAACCCAGGCGCCCAACCAACCAACCCAGGCCCAACTACACCC	590
OY		355	CCAATTTCAACAATTAATTTTTCAGCAACCCCAACCAATCATTAACCTGTGCAACCTCAACG	414
Db		591	AGGCGCCCACTACCACTCAGGCCCCCAACTACCAATCAACCAAGGCTGCACTACCCGGCGG	650
OY		415	CCATTTTCTCAACCAACCTCAACCAAGTCCCGCCCAACAGC	452
Db		651	CAACTACCCCGGCGCACTACCCCGGCGGCGCAACTAC	688

```

1  RESULT 9
2  US-09-103-429A-2
3  : Sequence 2, Application US/09103429A
4  : Patent No. 6187558
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Granados, Robert R
9  :
10 : APPLICANT: Wang, Ping
11 :
12 : TITLE OF INVENTION: A NO. 6187558el Invertebrate Intestinal Mucin
13 :
14 : TITLE OF INVENTION: CNA and Related Products and Methods
15 :
16 : NUMBER OF SEQUENCES: 4
17 :
18 : CORRESPONDENCE ADDRESS:
19 :
20 : ADDRESSEE: Brown, Plinasti & Michaels, P.C.
21 :
22 : STREET: 118 No. 6187558th Tlaga
23 :
24 : City: Ithaca
25 :
26 : STATE: NY
27 :
28 : COUNTRY: USA
29 :
30 : ZIP: 14850
31 :
32 : COMPUTER READABLE FORM:

```



```

: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90017
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/452,872
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/147,777
: FILING DATE:
: APPLICATION NUMBER: 07/876,289
: FILING DATE: April 30, 1992
: APPLICATION NUMBER: Unassigned (204/144)
: FILING DATE: October 29, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Marburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 204/153
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ. ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10747 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-452-872-1

Query Match      14.8%; Score 70.6; DB 3; Length 10747;
Best Local Similarity 53.9%; Pred. No. 1.8e-08;
Matches 145; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

: 181 CCACCCCCCTACACCAACACCAATTTCCCAACCAACCAACCACTTTCTCGG 240
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1718 CCACCCACCCCAAGCAGCCTCCCAACCAACCAACCACTCGCAAAACCAAG 1659
: 241 CCCCAACACCAATTCCTCGGCAACCAACCAATTTCCCAAGCCCAAGACCAAT 300
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1658 CCACCAACCCCAATGCGACCTCGACCAACCAACCAACCAAGCCACCAACAAG 1599
: 301 CCCCAACCAATTCCTCGGCAACCAACCAATTTCCCAAGCCCAAGACCAAT 360
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1598 CCACCAACCAATTAACCAACCAACCAAGCACTACCACTCGCTTACCAACCT 1539
: 361 CAACAATTAATTTTCAGCAACCAACCAATATACCTGTCACCACTTACAGCCATT 420
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1538 ATGGAATGCTTTTACGCAACCAAGTTAACAAGCTCCGAATTCACCAACCA 1479
: 421 CCTCAACACCTCAACCAAGTCCCAACA 449
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1478 GCACCAAGCTACCAACCAACCAACCA 1450

RESULT 14
: PCT-US93-03985-1/c
: Sequence 1, Application PC/TUS9303985
: GENERAL INFORMATION:
: APPLICANT: Roop, Dennis R.
: APPLICANT: Rothnagel, Joseph A.
: APPLICANT: Greenhalgh, David A.
: APPLICANT: Yuspa, Stuart H.
: TITLE OF INVENTION: DEVELOPMENT OF A VECTOR TO TARGET GENE
: TITLE OF INVENTION: EXPRESSION TO THE EPIDERMIS OF TRANSGENIC ANIMALS
: NUMBER OF SEQUENCES: 5
```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fulbright & Jaworski
: STREET: 1301 McKinney, Suite 5100
: CITY: Houston
: STATE: Texas
: COUNTRY: U.S.A.
: ZIP: 77010-3095
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/03985
: FILING DATE: 19930428
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Paul, Thomas D.
: REGISTRATION NUMBER: 32,714
: REFERENCE/DOCKET NUMBER: D-5478
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713/651-5325
: TELEFAX: 713/651-5246
: TELEX: 762829
: INFORMATION FOR SEQ. ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10747 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: PCT-US93-03985-1

Query Match      14.8%; Score 70.6; DB 5; Length 10747;
Best Local Similarity 53.9%; Pred. No. 1.8e-08;
Matches 145; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

: 181 CCACCCCCCTACACCAACCAATTTCCCAACCAACCAACCACTTTCTCGG 240
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1718 CCACCCACCCCAAGCAGCCTCCCAACCAACCAACCACTGCGCAAAACCAAG 1659
: 241 CCCCAACACCAATTCCTCGGCAACCAACCAATTTCCCAAGCCCAAGACCAAT 300
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1658 CCACCAACCCCAATGCGACCTCGACCAACCAACCAACCAAGCCACCAACAAG 1599
: 301 CCCCAACCAATTCCTCGGCAACCAACCAATTTCCCAAGCCCAAGACCAAT 360
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1598 CCACCAACCAATTAACCAACCAACCAAGCACTACCACTCGCTTACCAACCT 1539
: 361 CAACAATTAATTTTCAGCAACCAACCAATATACCTGTCACCACTTACAGCCATT 420
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1538 ATGGAATGCTTTTACGCAACCAAGTTAACAAGCTCGGACTTCGAATTCACCAACCA 1479
: 421 CCTCAACACCTCAACCAAGTCCCAACA 449
:   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1478 GCACCAAGCTACCAACCAACCAACCA 1450

RESULT 15
: US-08-147-777-3/c
: Sequence 3, Application US/08147777
: Patent No. 5914265
: GENERAL INFORMATION:
: APPLICANT: Roop, Dennis R.
: APPLICANT: Rothnagel, Joseph A.
: APPLICANT: Greenhalgh, David A.
: APPLICANT: Yuspa, Stuart H.
: TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
```

ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777

FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 07/876,289
FILING DATE: April 30, 1992
APPLICATION NUMBER: Unassigned (204/144)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/153
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 24979 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-147-777-3

Query Match 14.8%; Score 70.6; DB 2; Length 24979;
Best Local Similarity 53.9%; Pred. No. 2.4e-08;
Matches 145; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 181 CCAACCCCTTACACCAACAACCAATTCCTCCACACCAACCAACCTCTTCTCGG 240
DB 15950 CCACCAACCCCAAGCACTCCACCAAAACCAACCAACCACTGCAAAACCAACCAAG 15891

QY 241 CCCCAACACCATTCCTGCAACCAACAACAACCAATTTCCCAAGCCCAAGAACCAATT 300
DB 15890 CCACCAACCCCAATGCACTCCCAAGCAACCAACCAACCAACCAACCAACCAAG 15831

QY 301 CCCCAACACCATTCCTGCAACCAACAACAACCAATTTCCCAAGCCCAAGAACCAATT 360
DB 15830 CCACCAACCACTATACCAACCAACAAGCACTTACCAAGCACTCTCTGACCACTT 15771

QY 361 CAACAATAATTTCAGCAACCAACCAACAATTCATCCCTGCAACCTCAACGACCAATT 420
DB 15770 ATGAGATGCTTTACTGCAACCAAGGTAAACAAGACTCGCACTTCCAAATTCACCAACA 15711

QY 421 CCTCAACAACCTCAACCAAGTCCCAACA 449
DB 15710 GCACCAAGCTACCAACCAACCAACA 15682.

Search completed: May 25, 2003, 20:18:30
Job time : 52.7974 secs

